THIS PHYSICAL WORLD

The quality of the materials used in the manufacture of this book is governed by continued postwar shortages.

The Companion Volume

THIS LIVING WORLD

by C. C Clark and R. H. Hall

New York University

FIRST EDITION

1940

This Physical Wo

A College Course in Science

C. C CLARK
Professor of General Science
New York University

C. A. JOHNSON
Instructor in General Science
New York University

L. M. COCKADAY
Lieut. Comdr., U. S. N., Naval Instructor in Physics
United States Naval Academy

FIRST EDITION
Fifth Impression

McGRAW-HILL BOOK COMPANY, Inc.
NEW YORK AND LONDON
1941

THIS PHYSICAL WORLD

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

THE material progress made by man during the twentieth century has brought him into close contact with the applications of many scientific developments. Some understanding of the fundamental principles involved in the functioning of the physical world about us has become, therefore, a requisite of a good education. There are diversified opinions as to how this understanding may best be achieved by students majoring in fields of learning other than those specifically related to the sciences. For the past fifteen years New York University has been offering to some of its students science courses organized and presented as a survey of the major principles of all the sciences, together with their most widely used applications. As a result of these efforts a large number of students have developed a considerable measure of interest in the broad field of science and have secured a reasonably accurate understanding of a body of pertinent information. The present text is in part the outcome of teaching these courses.

The aim of this book, a companion volume to "This Living World," is to present in a readable style an accurate discussion of some of our basic knowledge about the physical world. Particular emphasis has been placed on the branches of science that have led to many of man's useful developments. In selecting the subject material, the authors have kept clearly in mind the particular interests and probable needs of students who are not pursuing some science major. It is not to be inferred from the selection herein contained that all the essential principles of the physical sciences have been included and that those subject areas omitted are not fundamentally important. To acquire even a general understanding of all the important knowledge of these branches of science requires some years of application to their study. In science courses that are limited in scope and time many

important topics must of necessity remain untouched. It is of singular importance, therefore, that emphasis should be placed on those topics which are of significance in interpreting our present physical environment and in understanding the applications of science that have provided many of the conveniences of modern life. In the present book this emphasis has been attempted by treating only those physical phenomena that have most vitally affected our present civilization and by presenting them in as comprehensive a discussion as possible within the available space. The aim has been to do this in such manner that the student who takes no additional science courses may obtain some useful understanding of physical phenomena, and that the student who pursues additional scientific work may find the treatment here of a valuable exploratory and basic type.

A further aim of this book is to develop the concept that the physical universe is orderly in its behavior and that logical conclusions result from the operation of scientific laws and principles. From time to time, historical accounts of the details of discovery of a scientific law are given, or an extended discussion is made of a fundamental principle, in order to point out in specific instances that there is a relationship between cause and effect in physical phenomena. The reader may then exercise his own thinking in reaching conclusions regarding similar relationships in other topics. Since scientific phenomena do have logical explanations, it is felt that an appreciation of this fact is one of the justifications of a survey course in the physical sciences. If it can be generally understood that the operation of natural phenomena is always the result of some fundamental cause, then logical and rational thinking may replace superstition and prejudice.

In organizing the content material, the authors have taken care to preserve throughout the text a connected account of scientific principles and their relationships. The various topics are not treated as unrelated fragments of knowledge; rather they are developed so as to show their relationship to each other. For example, the fundamental laws of wave motion are first illustrated with facts about sound and later used to explain the phenomena of light, X rays, and radio. Similarly, atomic and molecular relationships introduced in the subjects of atoms and

chemistry illustrate principles that also explain facts about heat, electricity, and electronics. One of the objectives of a survey course is the correlation of the pertinent knowledge of the various sciences. In arranging the content of this text, the authors have kept in mind the object of presenting a logical and connected story of the physical world.

The first chapter of this book includes some explanation of the nature of scientific laws and of the scientific method and describes how these have been applied so as to produce many of our physical conveniences. Immediately following this chapter is a consideration of the stellar universe and the solar system, causes of the seasons, and methods of reckoning time. The nature and structure of atoms as the material substance of the physical world, together with their behavior in producing many fundamental types of chemical changes, are discussed in the next chapters. These are followed with a consideration of the laws of heat and their application to our industrial life. The next chapters give an account of wave motion and radiant energy; and the characteristics and uses of sound, light, ultraviolet, infrared, and X rays and gamma radiation and radio waves are discussed. The fundamentals of electricity and free electron behavior and their extensive applications in modern society are next considered. The book ends with a chapter on methods of electrical communication, wherein the operating principles of the telegraph, telephone, and radio are explained in nontechnical language and the variety of their uses is set forth.

An annotated list of references for additional reading has been given at the end of each chapter for those whose interests may extend beyond the discussion of this text. In selecting references, the authors have included some popularly written books and magazines that are suitable for general reading and have listed some more technical books and professional journals for the specific and detailed information that they contain.

Acknowledgment is made of the help and cooperation the authors have received from a number of persons during the preparation of this volume. Thanks are expressed to colleagues in the general science courses of the School of Commerce, Accounts and Finance, New York University, for help in revising and in teaching the courses which led to the writing of

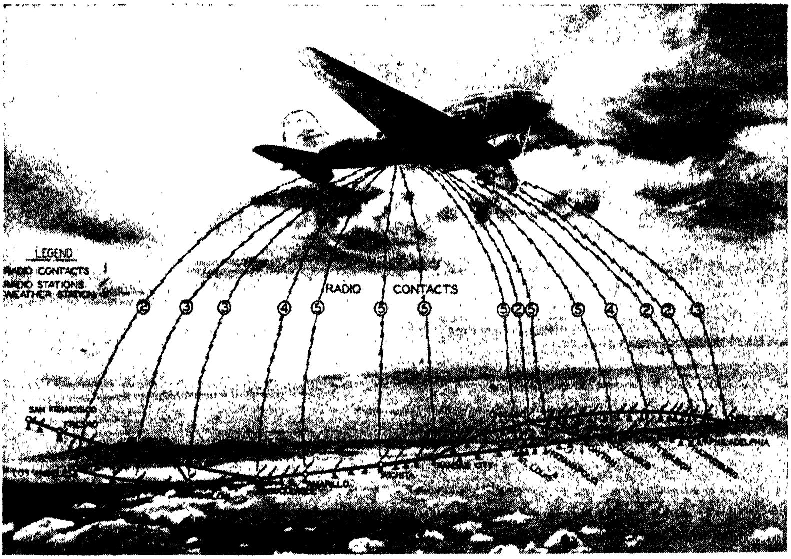
this manuscript. Particular acknowledgment is made to Dr. J. David Evans and Mr. T. J. Hanwick for their counseling on some of the chapters included in the book. The authors are especially obliged to the many persons and companies who have supplied them with photographs, specific credit for which is given where the pictures are used; to Mr. Theodore Miller for his cooperation and skill in making the drawings; and Dr. Clyde Fisher of the American Museum of Natural History, Mr. Robert D. Potter, Science Editor of *The American Weekly*, Profs. Lloyd B. Ham of the University of Arkansas and W. H. Hatcher of McGill University, and Drs. George H. Burnham of Hofstra College and H. C. Rentschler of Westinghouse Electric and Manufacturing Company for reading parts of the manuscript and offering timely and valuable suggestions for improvement.

C. C Clark,
C. A. Johnson,
L. M. Cockaday.

New York,
May, 1941.

CONTENTS

THIS PHYSICAL WORLD



T. W. A.

I: THE BEGINNING

Which Considers the Nature and Usefulness of Science

AN ECCENTRIC astronomer by the name of Tycho Brahe, who lived in Denmark during the latter half of the sixteenth century, spent a whole lifetime measuring accurately the positions of the planets as they revolved around the sun. At the time of his death in 1601 he had amassed several volumes of figures. For the first time in the history of mankind a mass of accurate data on the planets had been collected and recorded. From Brahe's calculations it was possible to check where the planets had been at any time within the period of his work. However, no interpretation had been made of these thousands of figures that would throw any light on how to calculate where the planets would be at some future date.

The figures were inherited by his pupil Johannes Kepler. After years of patient study of this large body of data, Kepler

discovered the relationships that existed between them. These relationships revealed the laws of planetary motion, which Kepler was able to state in three short sentences. Making use of the laws thus discovered, it is possible to calculate the past positions of the planets for any date and to predict their future movements. Such scientific and skillful study of nature had revealed one of its secrets and established a scientific law that has been of inestimable value.

On the other hand, man sometimes discovers a scientific law through misguided enthusiasm and by going contrary to the workings of nature. The discovery is attended by chagrin and discomfiture. One such instance concerns the transportation of the mongoose into the island of Jamaica.

The mongoose is a fierce little animal that inhabits a great part of India. Its natural enemies are snakes and rats, upon which it feeds. Years ago the mongoose was imported by the British into Jamaica to kill the rats that infested this island of the West Indies. The animal multiplied rapidly; the rats were soon killed; and the experiment was a great success. However, as soon as its natural food became scarce, the mongoose began to kill the chickens and birds: as the wild and domestic fowl were destroyed, the insects began to increase in destructive numbers. The island became infested with insects and mongooses. The latter are now greater pests than the rats, and there is no effective way to rid the island of them.

The balance of nature respecting living creatures is a definite natural law, and it had unknowingly been upset by man. As a result of this experience, the United States in 1902 passed a law forbidding importation of mongooses to this country, and they thus became one of our few outlawed classes of aliens. The fundamental laws regarding this balance in nature are now much better understood.

The foregoing illustrations are from two widely separated types of activity. One concerns the development of a law of astronomy; the other has to do with an illusive, although real, relationship in biology. In each case a discovery of a condition of nature had been made and a scientific law established. They are only two of many illustrations that show that the story of science embodies our understanding of the working of natural

phenomena. These conditions of nature may extend back into the past for hundreds of millions of years. They encompass most of our present activities and the processes of living. They often give us some insight into the happenings of the far distant future, and they invariably show a relationship existing between natural phenomena.

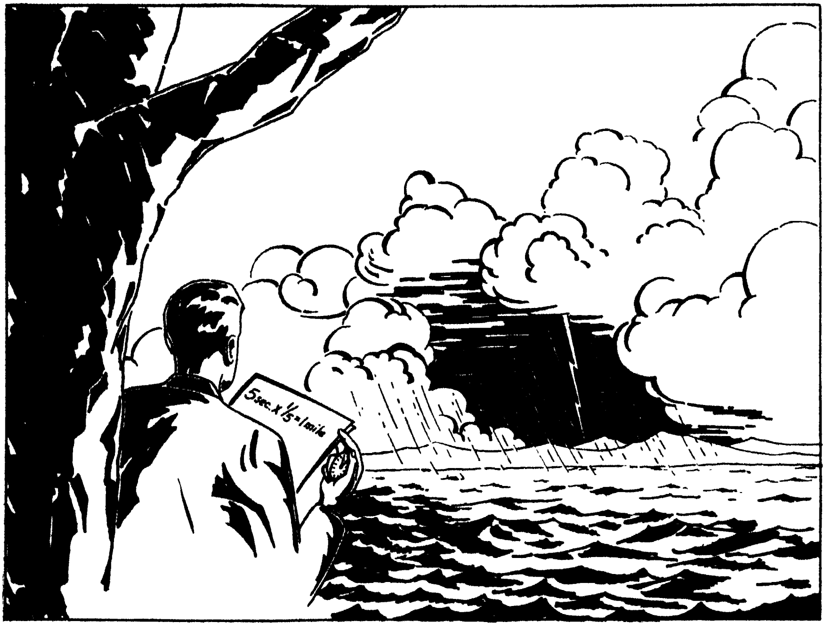
Laws Not Made by Man

Science, then, deals with the realities of nature as we are able to interpret them with the aid of our senses. In that respect it is a practical pursuit, for the mystical and magical lie outside its realms. This renders it at the same time more significant and more interesting, for fact is more engaging than fiction. The universe is orderly in its arrangement and behavior, and any sequence of natural happenings occurs in a definite and constant manner. A relationship always exists between cause and effect. As these relationships are known and understood, it is possible to state them as fundamental laws of nature.

In the course of thousands of years mankind has accumulated a large body of information regarding the make-up and processes of the physical universe. For a long time man has been seeking to discover the relationships that exist in nature and to describe the universe. In a large measure he has succeeded; but there is much left to be done. Much of the knowledge has been carefully and painstakingly classified and analyzed, and many of the fundamental relationships of nature have been deduced. The expressions of these relationships constitute the scientific laws.

An understanding of these fundamental laws of science enables anyone to comprehend the multitude of phenomena and happenings that affect his life continually. Also, in a great many instances, these scientific principles have been applied in such manner that we have been able to change or control our environment to provide most of the conveniences of modern life. These conveniences have not only enhanced our physical welfare but also allowed for the development of our intellectual powers and even influenced our ways of thinking.

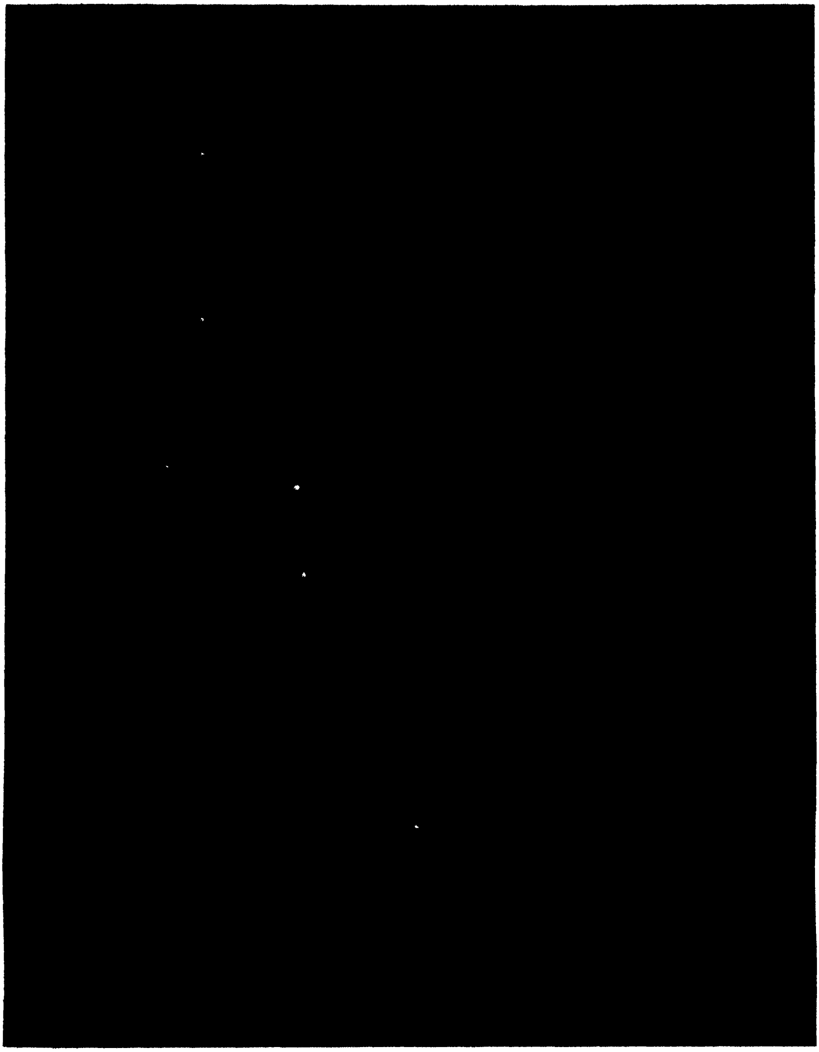
Some of the laws of nature are obvious and can be observed qualitatively by anyone. Such, for example, is the law of gravity. The proverbial apple is generally known always to fall to the



The number of seconds between a flash of lightning and the peal of thunder is a measure of distance of the storm.

earth when it breaks away from the parent tree. We know that it is pulled to the earth by the force of gravity, a force that moves the lighter apple to the heavier earth. This same force causes the earth to pull to its surface all objects lighter than itself, unless the lighter objects are buoyed up by a force greater than the earth's gravity.

Likewise, the approximate velocity of sound in air is known to most people. Who has not experienced the satisfaction of guessing the distance of a summer thunderstorm by counting the number of seconds between the flash of lightning and the peal of thunder? This number divided by five tells one the approximate number of miles from the storm center. He may know, for example, that sound travels about 1,100 feet, or one-fifth of a mile, per second. Therefore, in each five seconds' time the sound of the thunder will travel approximately one mile, and thus it becomes a measure of the distance of the thunder. The burning of coal or wood in the friendly fireplace is known to most people to be a chemical change taking place between the



Ladder of planets in the western sky as photographed by I. M. Levitt, of the Fels Planetarium Staff, Philadelphia, on Feb. 25, 1940, at 6:30 P.M. The positions of Mars, Saturn, Venus, Jupiter, and Mercury (reading downward) in their respective orbits were such as to produce this interesting spectacle of a nearly straight line as viewed from the earth. These five planets group themselves in similar manner as morning or evening stars approximately every twenty years. They were so visible in 1901 and 1921, and their laws of motion are so well known that it is possible to predict that they will again form a similar grouping in July, 1957.

carbon and the hydrogen of the wood and the oxygen of the air, with the liberation of a considerable amount of heat and light energy.

However, many of the laws of nature are very obscure and difficult to establish. They require the most precise instruments of measurement and the most abstruse mathematics. The idea of the emission of light by a glowing object in small, separate packets, or quanta, rather than in continuous fashion illustrates the point in question. The nature of this emission is obvious to no one looking at an electric lamp or the twinkling stars of a June evening. It becomes evident only when the most exact measurements and mathematics are applied by a highly trained observer.

Such illustrations as those just mentioned are typical of the kind of scientific laws that apply to specific and individual happenings. Everyone knows that if he jumps from the seventh-story window of a building or "bails out" of a soaring airplane he will immediately fall to the earth, and precisely this same thing will happen every time the "experiment" is attempted. Also, each time the outmoded kerosene lamp or the modern electric bulb is lighted, small quanta of radiant energy stream forth as visible light. Here it may be said that a given result must of necessity follow when a definite set of conditions or causes is fulfilled. These scientific laws applying to specific and individual happenings help to explain a multitude of natural phenomena that are continually being observed. They also enable one to predict with certainty what will happen in any instance in the future when a given set of conditions is known to exist.

In addition to the natural laws, in which a specific result must of necessity follow an applied set of conditions, there is another type of known relationships that is extremely useful in explaining many natural phenomena. It includes phenomena in the occurrence of which an element of "chance" is involved. The relationships that involve chance factors constitute the laws of probability. A little insight into the operation of these laws may be gained by a general consideration of a few of their simplest principles.

The popular concept of the word "chance" implies that it relates to a condition in which it is not possible to know or

control all factors influencing the outcome. An example may serve to make its specific meaning a little clearer. Suppose that concealed in an urn is an equal number of red and yellow balls that are exactly alike in every respect except color. If a person draws one of them blindly, it is certain that it will be either one or the other color. But whether it will be a red ball or a yellow one cannot be predicted, because that depends upon chance. In the process of drawing a ball, certain things are known; namely, the balls of the two colors are present in equal numbers, they are distributed at random in the urn, and they are identical in all respects except color. In addition to these known conditions other unknown factors are influencing the drawing. The respective positions of the red and yellow balls are unknown. Then there are certain factors that are an "act of will," which cannot be determined, for example the exact motion of the muscles in directing the fingers to a given spot and the complexity of these movements, which defies any possibility of foreseeing the final result.

The degree to which the element of chance exists varies in different situations. In the example just cited, the chances of getting a red or a yellow ball are equal, and the probability of getting either color is one in two, or one-half, so long as the conditions remain the same. Should some blue balls that are identical to the others in every respect except color be added and distributed at random in the urn and in number equal to each of the other colors, the probability of drawing a given color would be one in three, or one-third. If it is unknown, however, what the relative numbers of different colors are, then judgment of the degree of chance of getting a given color must be reserved until more information is available regarding them.

It seems to be a desire of the public mind to assign numerical values to the probable outcome of situations that involve the element of chance even though in many instances the unknown factors are such that it is impossible to calculate such a numerical value. It is common practice to say that the odds are seven to five, for example, on the outcome of a baseball game or a presidential election. Although considerable information may be available about the condition and expertness of the players

on the two teams or the political leanings of the voting public, many factors are unknown. It is impossible, therefore, to calculate accurately any numerical value for the probability of the outcome. The extent to which one accepts in any material way such numerical probability is determined in large measure by his willingness to try his luck on the chance factors' having been fairly accurately estimated.

On the other hand, when all the unknown factors can be learned and the degree to which they affect the outcome determined, it is possible to use the laws of probabilities to calculate what the results will be when a large number of incidents have occurred and the conditions have not changed. For example, it is possible to calculate with considerable accuracy just what percentage of his money one will lose when one plays over an extended period games of chance involving the use of machines that require the combination of numbers or other factors before a pay-off is returned to the player. The chances of such numbers and combinations' occurring may be calculated upon the basis of the structure of the machine and by using the laws of probabilities as they apply to its mechanical structure and operation. One may predict how often the "jack pot" on the slot machine, for illustration, or any other combination may be secured with continued playing.

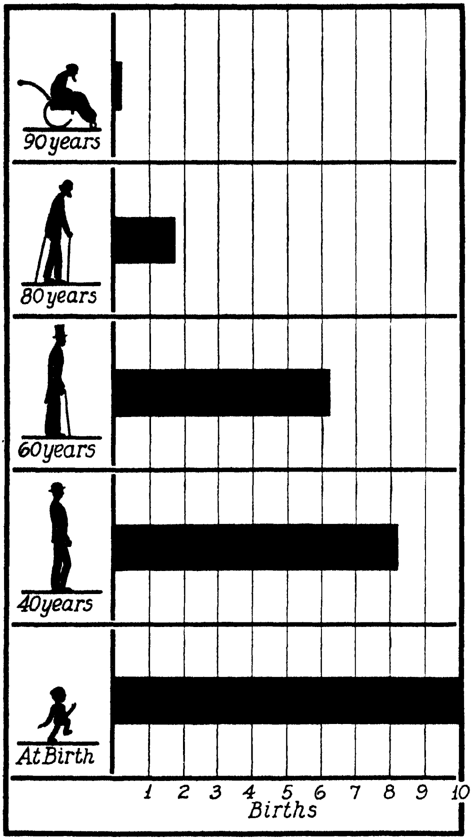
It is known that some slot machines will pay one jack pot per thousand plays over a large number of operations. This was determined by examining their mechanical construction, then applying the laws of probabilities. For example, in most slot machines it is required that three bars in a row, one on each of the three wheels, be secured simultaneously in order to get the jack pot pay-off. On some machines there are ten active symbols on each wheel, any one of which may stop at the point of observation. Only one of these symbols is a bar. Therefore, the chances of the wheel's stopping with the bar at the visible point is one in ten when there has been no tampering with the machine. The same chances prevail with the other two wheels, each being one in ten. However, the chances of getting the bar on each of the three wheels simultaneously is the product of the chances of each one individually. This is, of course, one in ten multiplied by itself three times, or one in a thousand.

This mathematical chance has been verified by actually playing such a machine one hundred thousand times and keeping an accurate record of each play. This record showed that if one plays the slot machine regularly and enough times, he may know with a high degree of certainty that he will win the jack pot on the average of once in a thousand plays. Likewise, it is possible to calculate the chances of getting other winning combinations and thereby determine exactly what percentage of money will be returned by the machine with long, consistent playing. The construction and operation of all types of "amusement" machines involving an element of chance are such that the percentage of money returned is less than that put into them, frequently much less. However, the laws of averages give no hint as to what any individual play will return. A losing combination on one play does not mean that the next play will produce a winning combination.

The laws of probabilities have been of value in many studies relating to population. It is an observed fact that in a country with relatively stable living conditions and over a reasonable length of time the numbers of births, deaths, marriages, divorces, the incomes, the spending power, and many other factors fluctuate within certain ranges which can be rather accurately determined. These conditions have been the basis on which have been established many economic, social, and commercial enterprises. Within any given year it is not possible to determine exactly how many deaths, divorces, and so on, may occur because of the lack of knowledge regarding the nature or degree of some of the chance elements. However, taking into account the known factors, definite predictions of average conditions over a period of years may be made.

These laws have been extensively employed by insurance companies to determine to whom their policies should be issued and the rates or premiums on such policies. When one buys an insurance policy, he pays a yearly premium on it of such an amount as to enable the insurance company to repay his beneficiary a stated amount at his death and also allow the company only a fair profit, even though he lives an alloted number of years. The amount of this premium is calculated with a high degree of precision.

This means that the insurance companies know how long the average person of any given age will live. The insurance salesman is able to tell with considerable accuracy how long it will be before the average person of a certain age departs this life, but he cannot tell how long it will be before a certain individual will be "crossing the bar." It is this average length of life which is the basis on which the yearly rate of premium is calculated. The accuracy and usefulness of the laws of probabilities in this case are attested by the fact that millions of dollars to beneficiaries are paid out annually, yet at the same time the insurance companies have become one of the largest and most stable financial institutions of mankind's entire business history.



Representing the average number of people who will live a given number of years after birth.

The laws of probability have been used extensively in the study of a wide range of scientific problems. They have been instrumental in providing much of our understanding of the size and speed of molecules in gases and the pressures created thereby, of the structure of atoms, of hereditary traits in man and other living organisms, and of a host of other hypotheses.

By the use of all scientific relationships and laws that man has discovered, the present understanding and description of the universe is accomplished. Thereby the knowledge that the

human race has accumulated is systematized and simplified so that it may be effectively used; also, such laws make it possible to look into the immediate or remote future in many instances and to predict with varying degrees of completeness what will happen.

The Scientific Method

During the long period of time that has been spent in the accumulation of a vast body of knowledge and in formulating a multitude of scientific laws, humanity has come into the possession of something more significant than a body of data and a system of classification. A method far superior to all others for the increase of our knowledge has been slowly evolved. This is usually referred to as the method of science, or the scientific method. The beginning of the modern scientific age is often said to coincide with the time when a large number of people began consciously to use the scientific method. Rapid increase in scientific discovery during the last few generations is in large measure the result of the application of this method on a wide scale.

There is nothing mystical or sacred about the scientific method. It is the "common-sense" way of securing information and of solving problems, and its processes now characterize the thinking of most normal persons. It is based upon observation and experiment, and it makes use of reason and induction in arriving at an understanding of the relationship between cause and effect or between forces and conditions applied and the results obtained.

This is a long step from the fundamental types of mental processes that characterized man's intelligent thinking for two thousand years from the time of Aristotle to Galileo. Much of the learning of the early Greeks was based upon logic and reasoning from assumed premises. This was characteristically a process of deduction, in which a fundamental premise was assumed to be true and from this premise the deduction made that certain details or specific situations and happenings must of necessity follow. This type of thinking took its cue from processes that had been developed in mathematics and successfully used in that field.

Aristotle applied the logic of mathematics to the theory of science, particularly to the physical sciences. He held that the essence of matter and of the universe was to be found in its "qualities," and he imagined that certain fundamental qualities (or premises) were combined in varying proportions to build up nature. From these he deduced by his logic the answers to many questions regarding the physical world. For example, he held that it was a fundamental quality of heavy and light bodies that the one would sink whereas the other would rise. He argued, therefore (and properly), that in a vacuum all bodies both heavy and light must fall with equal velocity. Such a condition he considered inconceivable, however, since it was incompatible with the assumed fundamental quality of heavy objects sinking and light objects rising. Accordingly, he concluded (incorrectly) that no vacuum could exist, and the question regarding a vacuum was thereby solved so far as he was concerned. He did not consider it necessary, in order to determine whether or not a vacuum could exist, to conduct an experiment that would also test and verify, or disprove, his fundamental premise about heavy and light objects.

Aristotle also held that one of the fundamental qualities of nature was that for bodies to remain in motion a force must be continually applied to them. Hence he reasoned and taught that to keep the heavens revolving an Unmoved Mover must continually be at work and, likewise, that for any object on the earth or elsewhere to remain in motion some outside force must be continually applied. He strongly rejected the idea of some of his contemporaries that bodies once set in motion would continue that motion unless opposed. There the matter rested for two thousand years until Galileo rolled some little balls down a nearly frictionless plane and found that they ascended a similar plane to an equal height owing to their own motion and thereby showed that objects in motion continue in that motion unless opposed by another force. This was exactly the opposite condition to that which Aristotle had reasoned. He had postulated a false premise and, of course, had reached a false conclusion as a result of his deductive reasoning.

It is not to be implied from these two examples that all Aristotle's premises and deductions were defective. Just the

opposite was true in many instances, for he made an exceedingly wide variety of observations and described many of the phenomena of nature with a high degree of accuracy. However, he did pass on to succeeding generations his deductive theory of science, and less capable minds made many incorrect deductions from assumed premises that were often not only unreal but absurd. Then during the Dark Ages even Aristotle's teachings were forgotten, and it became the mode of thought to accept all things upon authority or from some type of revelation.

With the beginning of the Renaissance the teachings of the early Greeks and other learned peoples began to be rediscovered, and the language of ancient philosophy and science became familiar to Western scholars after a lapse of hundreds of years. More important, a spirit of free inquiry began to develop. Not only did the scholars read Plato and Aristotle, but the idea was developed that knowledge unknown to the early Greeks might be discovered by observation, experiment, and inductive reasoning. Since the beginnings of experimental scientific research inaugurated by Leonardo da Vinci, Galileo, Newton, and Kepler mankind has come more and more to use the technique of investigation and reasoning, based upon existing knowledge and his discoveries, to further his understanding of the physical world. An appreciation of the scientific method and a general knowledge of how to apply it are becoming requisites for any individual who desires to progress in this increasingly complex civilization.

The scientific method is not a one, two, three process. It is characterized by familiarity with the existing knowledge of the problem being investigated, by open-mindedness, by accuracy of observation and analysis, and by honesty in evaluating the information gained. However, for purposes of understanding its nature, it may be broken down into about four steps.

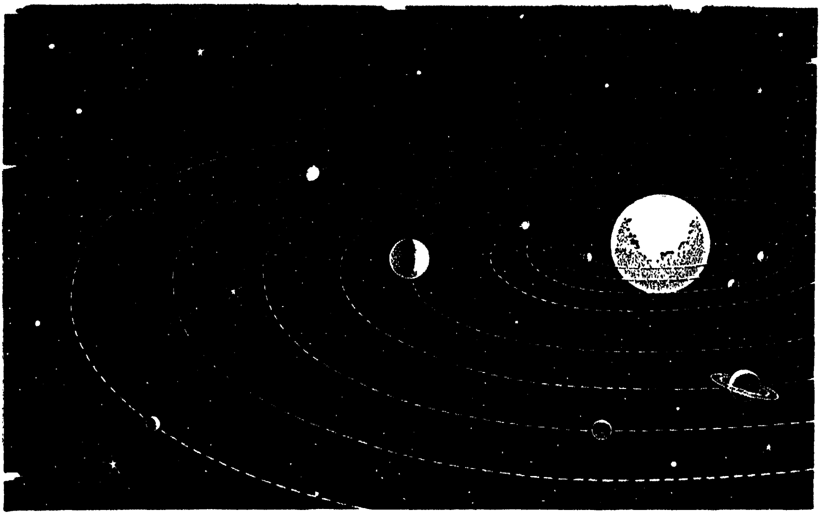
The first of these is the securing of data on the problem concerned. This may be done by careful observation, by collecting the results of previous experience, or by experimental techniques. In observation, information is gathered as it may appear, and incidents are noted as they occur. In this case it is essential that all happenings concerning the problem be observed or at least a representative sampling made. None is to be omitted

or overlooked because it does not fit into a preconceived theory. Such investigations may extend over long periods of time if the happenings occur infrequently. Or they may have to be made very quickly if the observed phenomena recur rapidly, as they do in many instances of nature.

Oftentimes much essential data on a problem may be secured by collecting the results of previous research. These data may exist in separated and isolated units, or they may be a part of another problem that is related to the situation being investigated. Where they have been secured under careful and accurate conditions, they may be of great value or save much time in the present problem. Under all circumstances the investigator should be familiar with the literature pertaining to the problem in hand and with the essential available information recorded there. Cicero, in "Orator," 34,120, said, "Not to know what happened before one was born is always to be a child." It is without doubt true that any person wishing to investigate or solve a problem is severely handicapped by not having the pertinent information relative to it that has been secured by previous efforts. Such information serves as a reasonable and rational basis from which to begin further investigation.

The term "experimental technique" usually refers to laboratory conditions where the happenings are made to occur in order that they may be observed. The situation is controlled so that the effects of different forces or conditions on the phenomenon being studied may be accurately determined while all other phenomena are kept constant in order to have only one variable present at a time.

For example, a radio-tube manufacturing company wishes to develop a new tube that will be most effective for the uses to which it is to be put. There are several factors to be studied, and each one must be investigated separately. Suppose that one of these is the exact composition of the metal filament and that other factors are such things as voltages applied, temperature, and spacing of different parts of the tube. Different materials for the filament are tried out, the other conditions of voltage, spacing, and temperature being kept constant, and exact data are recorded of the results obtained for each material tested. A few of the best are selected, and each one is separately tested



The laws of planetary motion were discovered by the observation of many thousands of different positions of the planets throughout the years, and inducing from this data the laws. These laws have been verified by repeated observations.

for its response when the voltages are changed. Likewise, each one is tested for its behavior under different temperature changes and also under different conditions of spacing of other parts of the tube. Finally, from the data obtained, it will be possible to select one particular filament material that will operate to best advantages at a specific and definite temperature and voltage and under definite arrangement of different elements of the tube.

When it is possible to reduce any investigation to the laboratory technique, the exactness as well as the ease of handling it is often increased. This technique has been applied to a great variety of scientific phenomena, and it has been in large measure responsible for much of our understanding of nature as well as the development of many of our material advantages.

The second step in the scientific method consists in classifying and interpreting the data that have been secured. The data on a given problem may be very large in number. Some are of no consequence except as duplications of other data. Some may be closely related; others, only remotely related. Therefore, it is essential that they be arranged in order or classified.

Classification will make it possible to discover the causal relationships that exist and to interpret their meaning. The discovery of the relationship between cause and effect in this manner is an inductive type of reasoning, and the process is sometimes referred to as induction. The inductive process may result in a generalization or the statement of a law of nature. It was in such manner that the laws of gravitation, of planetary motion, and of organic evolution were formulated. They are generalizations arrived at after observation, experimental determination, and classification made by many people of thousands of happenings over a long period of time.

After the generalization has been made or a law or theory formulated, it is necessary to test the law or theory. An effective way to accomplish this is to determine whether or not all specific incidents to which the law applies conform to the generalization. The idea of such testing of a scientific theory is based upon the fact that nature is orderly and not haphazard in its processes; cause and effect are related; things do not happen by mere chance. A specific result must of necessity follow a given set of causal circumstances. Once a generalization or law has been discovered, it is possible to go from the generalization back to an individual case. It may be reasoned deductively from the law that specific happenings must occur under a given set of conditions. It is, therefore, possible to predict what will happen in each instance when the causal conditions exist. This step in the scientific method is usually referred to as deduction. For example, if the law of gravitation is true, it may be deduced that all objects regardless of size or mass will fall to the earth with exactly the same speed in a vacuum. The question is, then, does this happen?

The fourth and final step follows immediately. The individual incidents derived by deduction from the generalization must be verified by repeated experiments or observation. If all such instances do conform, then the law is said to be universal. Repeated experiments with falling bodies have shown that all objects do fall to the earth with the same velocity in a vacuum. By many such tests the law of gravitation has been verified, and, furthermore, it has been established as universal in that it applies to all things. If some specific instances do not conform,

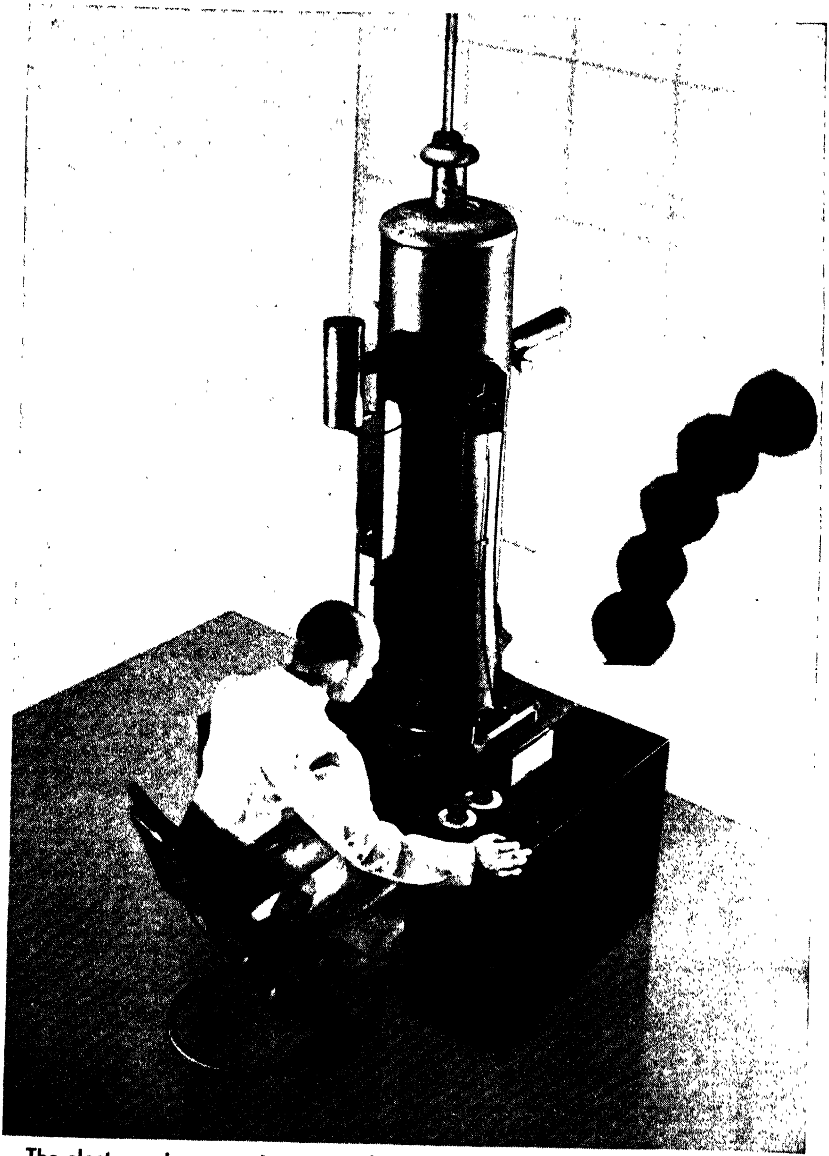
then the newly discovered law is true only under certain conditions. For example, the inherited characteristics of animals sometimes change suddenly. In biology these sudden changes are called "mutations." Mutations may be produced by exposing the reproductive cells to X rays. They may also occur under certain natural conditions, but they do not occur generally; therefore, mutation is a law of nature under certain conditions. The more accurately these conditions are determined and verified the more exact may be made the statement of the truth or law.

Such, then, is the general nature of the scientific method. The first two steps, gathering of data and induction, are essential in the formulation of any rational theory or generalization regarding the problem in hand. The last two, deduction and verification, are tests upon the extent of the truth of the conclusion or generalization reached. They subject to rigid tests the accuracy of any theory that may be postulated. As such, they constitute the essential difference between the modern scientific method and the older philosophical methods that were so long characteristic of man's thinking.

Applying the Foot Rule

It must be quite evident to everyone that the use of the scientific method involves quantitative measurement. In fact the understanding that man has of the make-up of the universe and of the behavior of nature is directly proportional to his ability to analyze and to measure such phenomena and happenings. Measurement is therefore an essential process of all scientific work, and some understanding of its significance is important in any study of science. Its value in a great many other activities of modern life is not less important.

For example, it is not to be wondered at that man remained ignorant of the nature of the sun, moon, and stars and the size and shape of the earth for millions of years when he had no methods of measuring their composition and movements. When Galileo constructed his first crude telescope, not only could his astonished friends look far out to sea at ships invisible to the unaided eye, but also he himself was able to view the sun and planets with sufficient clearness to determine something of their



The electron microscope is a recent development which is of outstanding importance in every field of science. It magnifies objects by electronic means twenty to fifty times as much as is possible with the finest optical microscopes in existence, and shows promise of extending the boundaries of knowledge in the study of human disease and other fields of science. Dr. Ladislaus Marton, inventor of the microscope, is seen at the controls, while the insert photograph shows streptococcus bacteria magnified 25,000 diameters. (R. C. A. photographs.)

size and movements. He raised the curtain on a new stage of human understanding (even though he was later cast into prison for it). The large telescopes of today increase man's vision a millionfold. They permit of such accuracy of observation and measurement as to enable him to determine the composition of stars billions of miles away or to predict the time of an eclipse of the sun years in advance with an error of only a few seconds. In a somewhat more practical sense, the excellence of the modern motor car, which operates so smoothly that one is unaware of machinery, is possible because of the fine precision used in its construction. One illustration is that the piston rings in the cylinders fit to an accuracy within thousandths of an inch.

Many phenomena are now subject to measurement, while some still lie outside this realm. Such factors as distance, composition of matter, time and space, electric phenomena, and radiant energy may now be measured. Man can determine the size of atoms or stellar distances. He analyzes the products of complex chemical reactions and understands the chemical composition of living tissue. He measures minute electric currents emitted by a tiny photoelectric cell or the electric output of huge dynamos. He can measure the wave lengths of X rays in millionths of an inch or the velocity of light in hundreds of thousands of miles per second. Such things constitute established knowledge.

Such factors as the exact nature of electricity and of the energy that is the "spark of life" are not subject to known methods of analysis and measurement. Electricity is known to consist of electrons and positrons, *i.e.*, negative and positive charges, but what such charges are has not yet been determined. What happens to life after death of the physical body cannot be detected or measured. Whether or not human mental processes may project themselves through time and space except by use of the physical senses is as much a speculation today as in ages gone by. Such things lie outside the realm of measurement, and man possesses no knowledge of their fundamental nature.

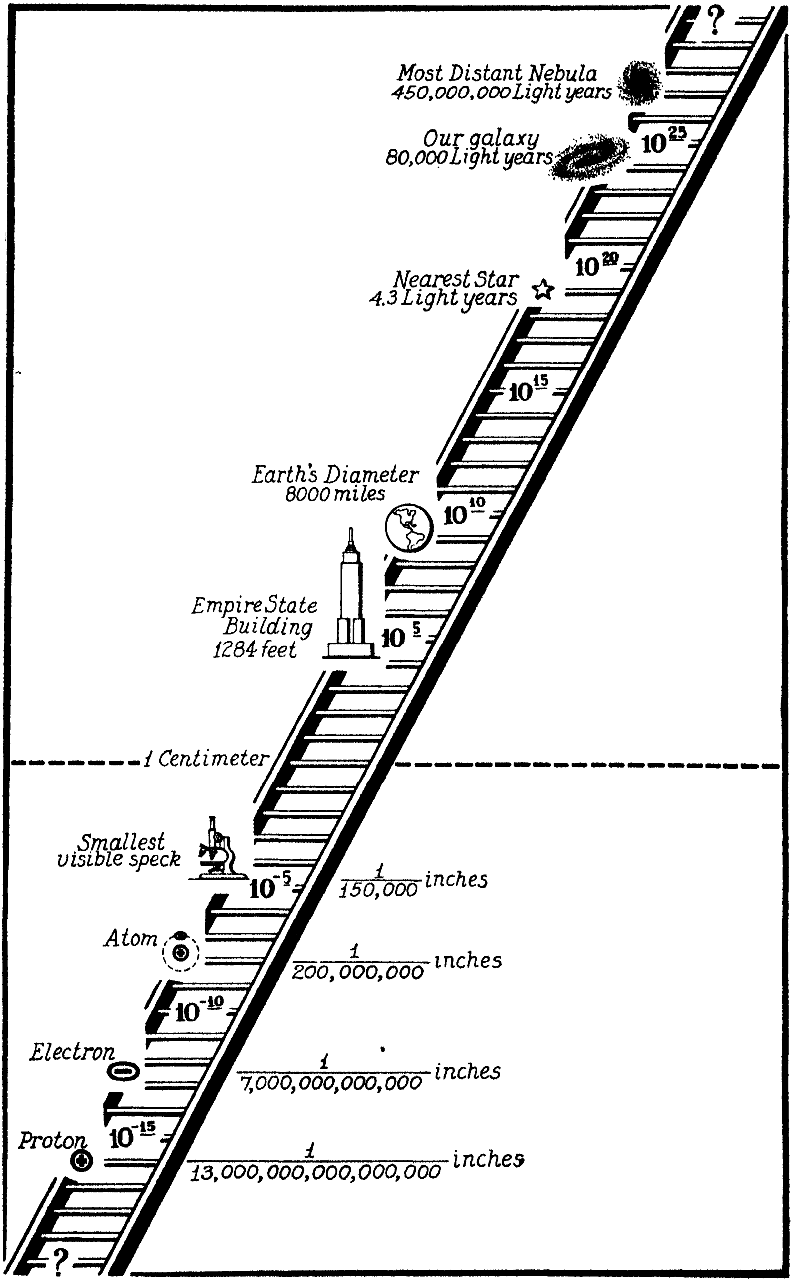
Big and Little Things

The application of measurement to such a variety of things has revealed an enormous range of size in materials of the known

universe. Everyone is familiar with the sizes of the common objects of the earth, such as mountains, houses, man himself, fruits, dust motes, even some microscopic particles. Few, however, have ever seen in clearly defined relief such large objects as the planets or distant stars or the extremely small bacteria that continuously infest their own bodies. But even these do not constitute the range limits in sizes of objects making up the universe. The larger stars and galaxies and their distances are of such magnitude as to lie beyond the limits of man's sensory comprehension. The minutely small objects such as molecules or electrons are outside human vision even with the use of the highest power microscopes.

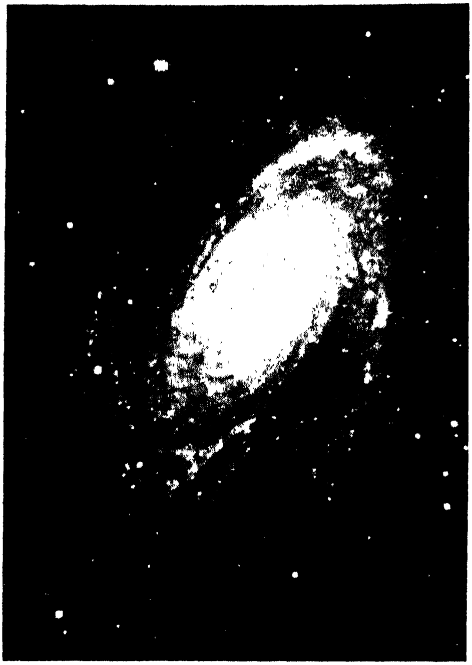
However, just because man cannot see them is no reason why extremely large and extremely small objects cannot exist. It is a tribute to precise measurements and mathematical deduction that his comprehension and understanding of the universe are what they are and have been extended manyfold beyond human vision.

The range of sizes to be found in the universe is shown in the accompanying chart. This is arranged in the form of a ladder in which each rung represents a multiple of ten of the one below it. The rungs near the center represent the sizes of average things known to man, such as dust particles, the average-sized man, the Empire State Building. The center rung represents an exact size of 1 centimeter, the centimeter being a unit of length in the metric system of measurement, and a little less than half an inch. The next rung above is 10 centimeters. The second one above is 10^2, or 100, centimeters; the third one 10^3, or 1,000, centimeters, and so on. The ninth rung represents 10^9 centimeters, or a distance about equal to the diameter of the earth. The nineteenth represents the distance to the nearest star, a distance so great that it is not calculated in centimeters or miles but in light-years. (A light-year is the distance that light will travel in one year while moving at the extraordinary velocity of about 186,000 miles per second.) To calculate the distance of light-years in miles it is necessary to multiply this figure by 3,600 in order to get the distance traveled in one hour. The product secured is multiplied by 24 to determine the distance for one day, and this in turn multiplied by 365 to secure the distance



Size ladder of the universe.

traveled in one year—a distance of over five trillion miles. Our nearest star neighbor is 4.3 light-years away, which is quite a sizable journey, even at the speed of light.



This spiral nebula in the constellation Ursa Major consists of millions of stars and is about two million light-years from the earth. (Science Service photograph.)

This is small distance, however, as compared with the size of our Galaxy of stars. The Galaxy in which we live is about 80,000 light-years across. Out beyond the Galaxy are other galaxies, each composed of thousands of stars. Some of the galaxies are at least 300 million light-years from the earth, and astronomers believe that even the most distant known ones do not represent the entire extent of the universe.

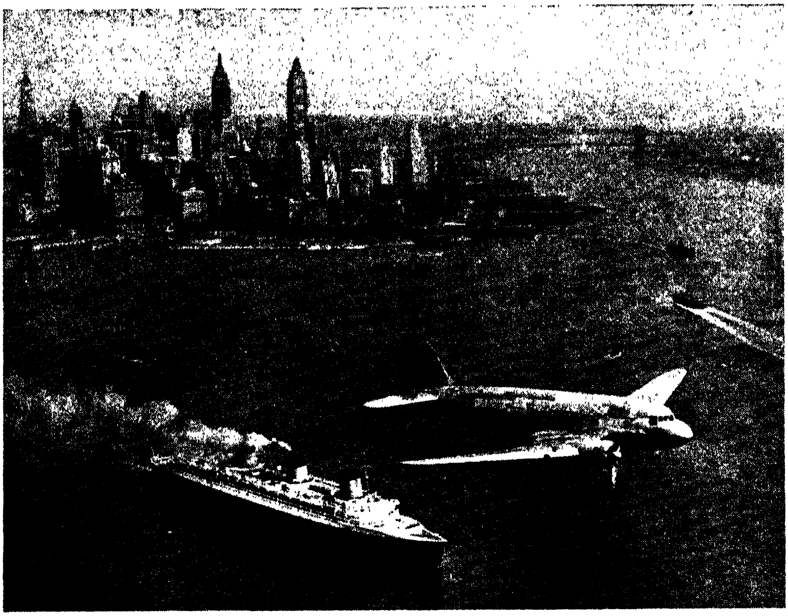
The other end of the ladder represents the smallest things known. The smallest speck that the unaided eye can detect is about one five-hundredth of an inch in diameter, represented by the second rung below. This is about one-half of one-hundredth of a centimeter, or 10^{-2} centimeter. With the highest power microscopes, objects of about $\frac{1}{150,000}$ inch, or about 10^{-5} centimeter, are visible. But these do not represent the limit. The atom is approximately 10^{-8} centimeter in diameter; the electron, about 10^{-13} centimeter; the diameter of the proton, probably about 10^{-16} centimeter. This means that it would take something like 13,000 billions of protons laid side by side to make an inch. There may be other particles unknown to man that are smaller still.

Both ends of the size ladder are still represented by question marks. They constitute areas of the unknown in man's knowledge. What lies there is surely much more challenging than what exists in the interior of African jungles. Should those who are reading this story live to "man's allotted three score and ten years," they might see the question marks pushed further back. In fact even now science gives us from time to time news of further insight into the unknown.

The Fuller Life

The discovery of fundamental scientific principles and the development of the scientific method constitute remarkable accomplishments; however, the most obvious and spectacular results of science are the material achievements, or scientific inventions. Scientific invention has completely transformed man's ways of living during the past century. His mode of travel, his means of communication, his ways of earning a living, the conveniences of his home life, the treatment of disease, the maintenance of body health, and his means of entertainment and amusement are almost completely the products of science. As a result, space is being annihilated, disease is being eliminated, material conveniences are increasing, even man's ways of thinking are changing.

In the field of communication the human voice may now be sent around the earth by telephone in a few minutes. This was first dramatically performed in San Francisco a few years ago at the meeting of the National Electric Light Association. The president of the association sent a greeting by telephone to one of the directors seated by his side in the same room. The message was sent across the Pacific, wired to London, relayed to New York, and finally returned to San Francisco. In a little over five minutes it was received in the director's room, and the speaker heard his own message after it had circled the globe. The radio has even surpassed the telephone. The coronation of King George VI in 1937 was listened to by peoples all over the earth. The king's message was heard at the most distant points only a fraction of a second after the words fell from his lips. At present, broadcasts from foreign lands are heard daily.



Much of the fuller life represented by this picture of lower Manhattan Island in New York City and the S.S. Normandie steaming out to sea has resulted from scientific research and engineering achievement. (Courtesy of Transcontinental and Western Airways.)

The first long step in this development came in 1844, when a young New York University professor, Samuel F. B. Morse, sent from the United States Supreme Courtroom in Washington to Baltimore, over a simple handmade telegraph laboriously designed, the first telegram—"What hath God wrought." The world-wide use of the telegraph at present is a monument to Morse's remarkable achievement, as the words of his first message were a mark of his humility. Our present-day, twenty-odd million telephones are all patterned after Alexander Graham Bell's original invention in 1876. Guglielmo Marconi sent the first radio message across the Atlantic in 1901, and the fundamental principles that he demonstrated then provide the radio of today.

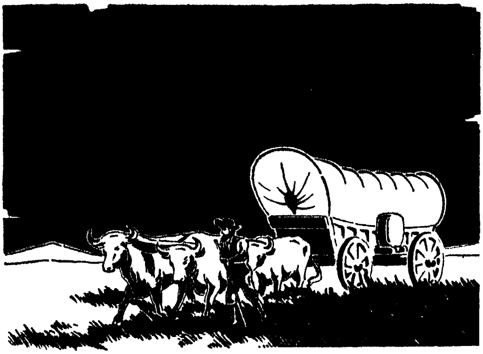
The use of these developments is so widespread that we now take them for granted. Such an attitude is not to be condemned; however, their existence today does not erase the slowness of man in developing them. In 490 B.C. the Athenians defeated the Persian invaders at Marathon, and the Olympic champion

Samuel F. B. Morse and his early telegraph. (Science Service photograph of original photograph on steel made from life by John Sartan.)

Pheidippides was dispatched to the capital with the news. He ran the twenty-two miles to Athens only to fall dead at the outskirts of the city as he gasped, "Rejoice, we conquer." For over 2,200 years no fundamental changes in methods of communication were made over this first marathon until Morse sent his first telegraphic message. Following this, the discovery of many scientific facts and laws had to be made before the development of modern radio was possible.

The same insight that discovered these laws now shows us many imperfections in our present inventions. There is yet plenty of room for advancement, even in methods of communication. The voice that you hear over the telephone still does not sound much like that of the person to whom you are talking. The reproduction at best is only second-rate, although the United States has the best telephone system in the world. Transoceanic telephoning is accomplished now only by using radio waves, and it is often wheezy and fading. An under-ocean

concentric electric cable would give better results; but the laying of such a cable has never been attempted.



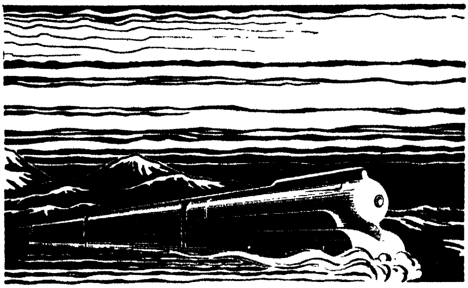
The covered wagon was the method of overland western travel until after the Civil War.

Likewise, commercial radio reception is still distorted and punctured by static. Theoretically as well as practically, static may be eliminated from the radio loudspeaker, but such perfection is yet to be achieved on a wide commercial scale. Television of the fine detail now scientifically possible has been seen by only a very few people, and how much longer we shall have to wait for general television no one knows. When it has been added to radio telephony, the last great frontier in communication will have been crossed. Pictures and sound may be sent at the speed of light, the highest attainable speed in the universe.

Transportation has shown remarkable development in the last century. In 1852 it took Ezra Meeker four months to travel in his caravan of ox-drawn covered wagons over the Oregon Trail from Omaha to Oregon. Now the Union Pacific streamlined trains make regular scheduled runs from Chicago to San Francisco in 39 hours. These trains are equipped with such scientific and engineering developments as to make travel on them, from the point of view of speed, safety, and comfort, a luxury undreamed of a few generations ago. Airplanes now fly on regular schedule from New York to Los Angeles in about 17 hours, carrying passengers, mail, and baggage. These planes are provided with many scientific control and indicator instruments. In 1935 the Pan American Airways inaugurated systematic airplane service between San Francisco and Manila, and regular transatlantic airplane service was begun in 1939. The planes on the Pacific route make their journey in 65 hours of actual flying time, and the Atlantic route planes have an actual flying schedule of 24 hours between New York and Portugal. In both cases, the planes are in constant touch with airports and weather-

bureau stations by means of wireless telephones, and often their direction of flight is guided by radio beams.

But with all this our transportation systems are not perfect. Much of the rolling stock of railroads is obsolete, slow, and inefficient. Furthermore, railroad companies still stop a 500-ton freight train manned by a crew of four or five men to pick up a few ten-gallon cans of milk. These are only a hint of the short-comings connected with rail transportation, but again it must be said that the United States has by far the most extensive and finest railroad system in the world.

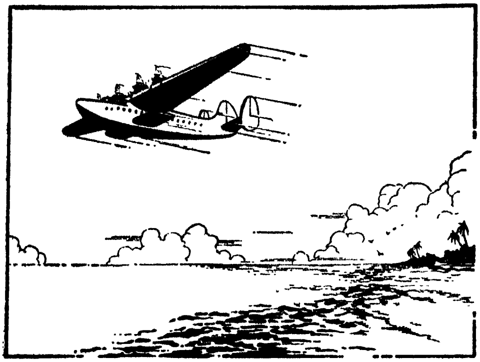


Streamlined trains now make the journey from New York to Chicago in 16 hours.

The automobile has literally put America on wheels, yet the modern motor car uses only about 8 per cent of the heat value of its gasoline fuel. Should even the unburned fuel in the form of carbon monoxide that goes out the exhaust pipe be completely used, the average mileage could be increased to about forty-five or fifty miles per gallon. But to retrieve this lost energy requires more knowledge of the chemistry of exploding gasoline than we now possess.

Man is now able to fly around the earth in less than four days without mishap, and commercial air records show over fifteen million passenger miles per death from accident. However, great transport planes do crash into mountains and dive into oceans, killing passengers and crew alike. The death rate in passenger planes is almost twice that of passenger automobiles and about fifty times as great as passenger deaths in railway accidents. Safety in airplane travel is still an important problem to be solved. Rapid advancements have been made in the last two or three years in perfecting the mechanical features and indicator and control devices of the planes. Pilot error is where the blame is often placed, but even including the errors of human judgment and human frailties, it is likely that the pilot is little at fault. Back of the pilot are weather and the powerful forces of nature

in the clouds. Although these adverse forces cannot be controlled by man, they need to be avoided. More accurate and immediate reports of weather conditions all along the route are one necessity in promoting safety in air travel.



American Airliners fly the Pacific in 65 hours.

The foregoing accounts are a few instances in which a concentrated and deliberate use of known scientific phenomena has been made in order to improve the lot of mankind. Here the inventions were mostly the results of attempting to find practical applications of scientific laws. Many great industrial laboratories are now organized for the explicit purposes of discovering more useful applications of science to the everyday affairs of life and of overcoming some of the present inadequacies of our material developments.

Unexpected Practical Worth of Scientific Research

Not all scientific inventions are the result of definite effort along such lines, however. In fact men doing research in pure science are often not concerned with its practical worth; their purpose is to discover fundamental truth regarding a phenomenon. Should this information later prove to be of commercial or economic value, it becomes a tribute to those who made the discovery, though it may not have been of particular concern to them at the time. The development of neon lighting and the widespread use of photoelectric cells are typical of this condition.

About seventy-five years ago an obscure German instrument maker had produced some small glass tubes that were closed at each end and from which the air was partly exhausted. A high-voltage electric charge was discharged through these tubes by means of electric contacts sealed in their ends. A peculiar glow of light was produced when the discharge of electricity occurred. A little later Sir William Crookes studied this

beautiful fluorescent display and discovered that it was produced by a stream of negatively charged electric particles passing through the tube. These streams of particles he called the cathode rays. For the next fifty years they proved to be a wonderful means of delving into the structure of matter, but during this time no commercial use was made of them.

Coincident with this development but entirely independent of it, another investigation was being made of the composition of the air. For nearly a hundred years it had been known that a little less than one per cent of the air could not be accounted for. This was usually explained on the basis of inaccuracies in measurements or an "elusive" something that could not be detected. From a practical standpoint this less-than-one per cent error in the analysis of air had no significance; but pure science does not share this view. In 1894 two eminent British scientists, Lord Rayleigh and Sir William Ramsay, discovered in the air traces of the gases argon and neon. Within four years they had added to the list two other inert gases, krypton and xenon, the four gases accounting for the missing one per cent of the air. This was a great discovery in the realm of pure science, but probably no one suspected at the time that it would have any industrial importance.

The development of the neon light came some twenty-five years later when pure neon gas was produced on a practical scale, and it was found that the beautiful glow discharge of the Crookes tube was associated with neon and the other rare gases of the air. From these two discoveries came the neon-lighting industry, now one of the country's great enterprises.

The photoelectric effect was first observed by the German scientist, Heinrich Hertz, in 1887 when he found that light waves falling upon a spark gap enabled an electric current to pass more readily across the gap. The following year a fellow countryman, Hallwachs, discovered that when light fell upon a zinc plate, the plate always became charged with positive electricity. He soon discovered that when light fell on such metals as zinc, sodium, and a few others, electrons, or negative electric charges, were ejected from the metal, thus leaving it positively charged. The number of electrons, *i.e.*, the size of the electric current, emitted by such metals when light or other sensitive radiation falls upon

"Sound-over-light" transmission demonstration by General Electric scientist at the New York World's Fair. (Courtesy of General Electric.)

them is directly proportional to the intensity of the radiation. By 1892 Hallwachs had used this photoelectric effect to measure the ultraviolet radiation from the sun.

There are few more interesting developments in all physics than the growth during the last forty years of the surface photoelectric effect from the obscure physical phenomenon discovered by Hertz and Hallwachs to one of profound theoretical significance and of surpassing commercial importance. The commercial uses of photoelectric cells have multiplied with bewildering rapidity during the last several years. They are used in a wide variety of measuring instruments where a more sensitive, more accurate, and more reliable instrument than the human eye is necessary. They are used in a great number of counting and control devices. In addition, they are an important factor in one of the country's great industries, the talking motion pictures, not to mention the transmission of telephotos and their use in television.

These brief accounts of scientific discovery and invention shaping man's life in the modern world could be multiplied almost indefinitely. Achievements in the study and forecasting of

The light waves are focused on a photoelectric cell where they are converted into electric signals which may be used to reproduce the original sounds, as demonstrated by General Electric scientist at the New York World's Fair. (Courtesy of General Electric.)

weather, in the analysis of protoplasm which is the physical basis of life, and in the synthetic products of the chemical laboratory are no less remarkable. The practical value of science is everywhere apparent. As Little in his "Contributions of Science to Manufacturing" says, "Science, in its industrial applications, is as intensely practical as a market report or balance sheet."[1]

REFERENCES FOR MORE EXTENDED READING

HASLETT, A. W.: "Unsolved Problems of Science," The Macmillan Company, New York, 1937.

Many of the most provocative problems that the scientists are now trying to solve are interestingly and searching discussed in this book

LEVY, HYMAN: "Modern Science: A Study of Physical Science in the World of Today," Alfred A. Knopf, Inc., New York, 1939.

The person wishing to obtain an analysis of the methods, achievements, and effects of science may well afford to read this excellent book by a leading British scientist.

HOGBEN, LANCELOT T.: "Retreat from Reason," Random House, Inc., New York, 1937.

[1] LITTLE, ARTHUR D., "The Contributions of Science to Manufacturing," *Annals of the American Academy of Political and Social Science*, May, 1925.

The author, an authority on many scientific subjects, has produced here a little book in a delightful style wherein he laments the established order of things and emphasizes the rationalist standpoint of scientific research into the fundamental human needs.

FREEMAN, IRA M.: "Invitation to Experiment," E. P. Dutton & Company, Inc., New York, 1940.

Many experiments in physics are discussed in such a simplified manner that the interested layman may perform them outside the classroom and thereby gain some understanding of the physical world.

MITCHELL, S. A.: "Eclipses of the Sun," 4th ed., Columbia University Press, New York, 1935.

This book contains a well-written discussion of past eclipses as well as an excellent presentation of modern methods of scientifically observing them and a thorough account of the facts and theories resulting from their study.

WENDT, GERALD, "Science in the World of Tomorrow," W. W. Norton & Company, Inc., New York, 1939.

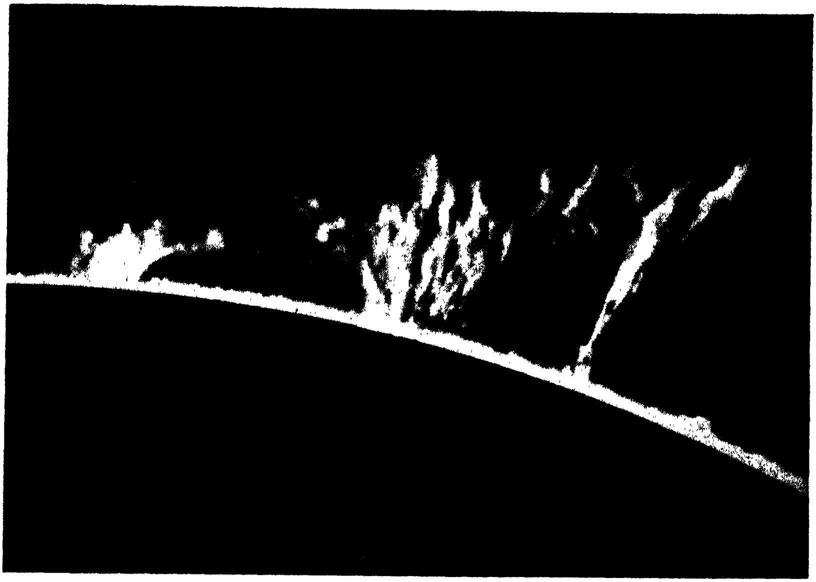
An interestingly written book in which is set forth some advances in the use of natural resources, in transportation, in communication, and in synthetic materials, and in which it is pointed out that development of the scientific attitude is one of the most important contributions of science to mankind.

Scientific Monthly, published by The Science Press, Lancaster, Pa. for A.A.A.S.

An illustrated magazine published for The American Association for the Advancement of Science and devoted to the general diffusion of science knowledge.

Journal of Applied Physics, published by The American Institute of Physics, New York.

The articles appearing in this journal are written in a considerably advanced technical style and relate to applied physics and original research

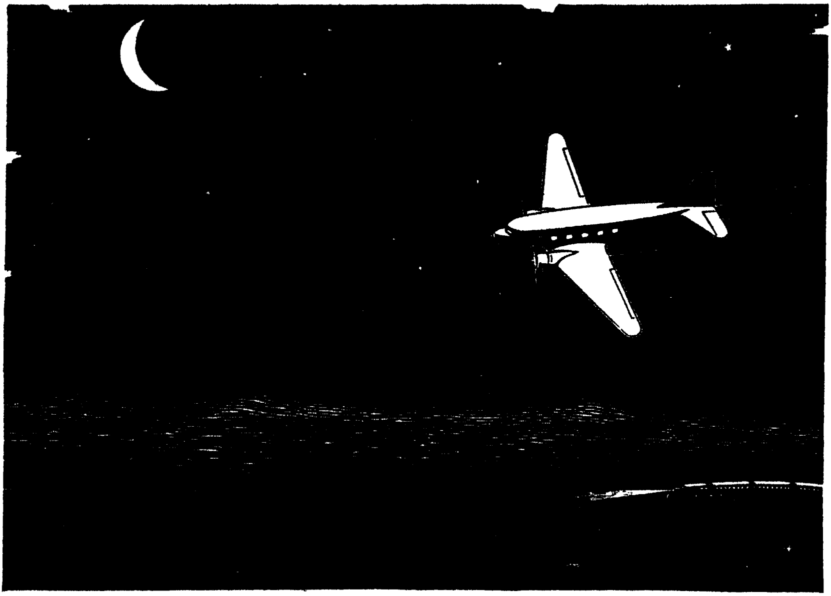


Science Service.

2: LOOKING UPWARD

To See the Sun and the Stellar Universe

THE earliest recorded eclipse of the sun is said to have occurred on Oct. 22, 2137 B.C. Over four thousand years ago —about fourteen hundred years earlier than the recording of eclipses by any person of another nation—two Chinese royal astronomers, Hsi and Ho, made a record of the time and circumstances of this spectacular phenomenon. But the two astronomers were not faithful in fulfilling the requirements of their profession. Instead of predicting the exact day of the eclipse and preparing to shoot arrows and beat drums in order to deliver the sun from the monster that was devouring it, they had been imbibing too freely of the native wines and were experiencing a prolonged state of intoxication. Thus, great consternation and confusion reigned in the empire because the ceremonies were not carried out, and Chung K'ang, emperor in the Hsai dynasty, ordered the priests' heads chopped off.

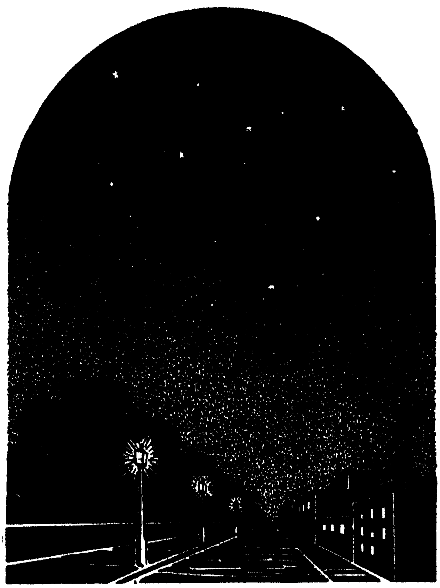


The tiny lights below may reveal to the air traveler many incidents occurring there.

The unhappy fate of Hsi and Ho not only has stimulated folklore and bursts of poetry but also is indicative of the change in our thinking regarding celestial happenings and the remarkable advance in our knowledge of the stellar universe. These astronomers began a series of observations that are now carried out with the most precise and careful techniques known to man. Even today astronomers travel to the remote ends of the earth to make careful, technical studies of the sun at times of total eclipse. From such studies they are able to arrive at a more definite understanding of the conditions in the sun and the make-up of the universe.

Observing the sun during an eclipse is only one of a great many types of astronomical investigation that have yielded knowledge regarding the universe. The discoveries made since ancient times have greatly changed the philosophy of most peoples and have also been an important factor in the development of science. For example, the commanding group of stars known as the constellation Orion was by legend a giant hunter who stood in the sky with a menacing club in one hand and a lion's skin in the other. Now we know that in this constellation

are two of the most remarkable bodies in the heavens. One of them is the reddish star Betelgeuse, and the other is the blue-white star Rigel. Betelgeuse is twenty-seven million times larger than the sun, and Rigel radiates about fifteen thousand times as much light as the sun. The petty imaginings of the ancients have been replaced with magnificent truth. There is still mystery in observing the heavens on a clear, dark night or in viewing a total eclipse of the sun. But this mystery is now fortified with a knowledge of many remarkable things that the stars and sun have to tell.

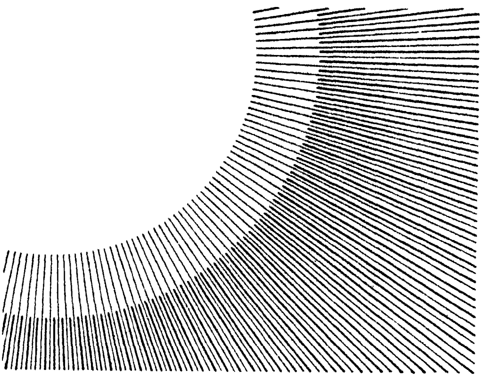


The stars are not tiny specks inside a hollow dome.

Messages from Light

A recent public interview with a stewardess on a transcontinental airplane revealed that she liked most to fly at night. As the plane soared along on its journey, the passengers occasionally glanced at the tiny lights on the landscape below. Some would say, "How pretty," and turn back to their magazines or doze off to sleep. But she would look a little longer, for all these lights revealed to her imagination many happenings and many activities of people below. A glance at the stars in a cloudless night sky may bring forth the statement of "How pretty," or even hold one spellbound in awe and wonder. But the light from these twinkling stars has revealed much information of what is happening out in celestial space. Man has built up a remarkable story about the sources of this light. With large telescopes he has concentrated the light and accurately measured the number, sizes, and positions of the stars. With precision spectroscopes and cameras he has determined their composition, temperatures, and velocities. Perhaps the most remarkable revelations that

the telescope and spectroscope have brought us concern the true nature of the sun and stars.



The stars are not, as the ancients believed, tiny specks inside a hollow dome. They are not like so many lights on a ceiling, to be turned out each morning and lighted again at dusk. They are not the gentle, twinkling lights that to the unaided eye they appear to be. Rather they are large, surging, tumultuous masses which shine by their own light and are comparable in general with the sun. This light must be enormous if we see it at all, as the stars are so distant from the earth. The stars are suns, similar in general character to our own sun. As a matter of fact, our sun is an ordinary star; but because it is so much nearer, it does not look like one. Many of the stars are much larger than our sun, whereas a great number are smaller; many are much hotter and brighter; some are cooler and darker.

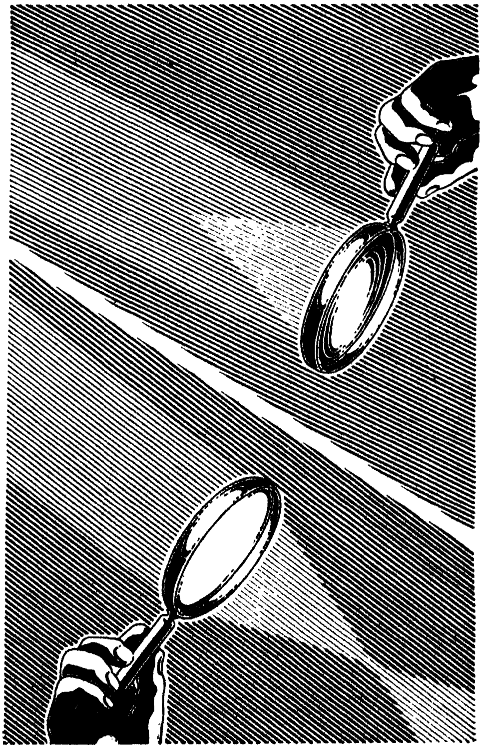
Substance of Sun and Stars

Of what, then, are these shining objects of the sky composed? When man wishes to determine the composition of an ore deposit, he digs into the earth and gets a sample; then in the chemical laboratory he analyzes it to determine its nature and value. Wishing to learn about the composition of the sun and stars, he can take from them only the light that comes streaming to the earth and analyze it. However, this method is now just as accurate and revealing as if he could reach out and get a handful of star dust or of the sun's substance and subject it to his searching gaze. He analyzes the light that comes from these heavenly bodies and by meticulous observation and study learns the story that it has to tell.

The first problem is to get enough light to analyze, particularly in the case of the less bright stars. The human eye is so small that it can intercept only a very narrow

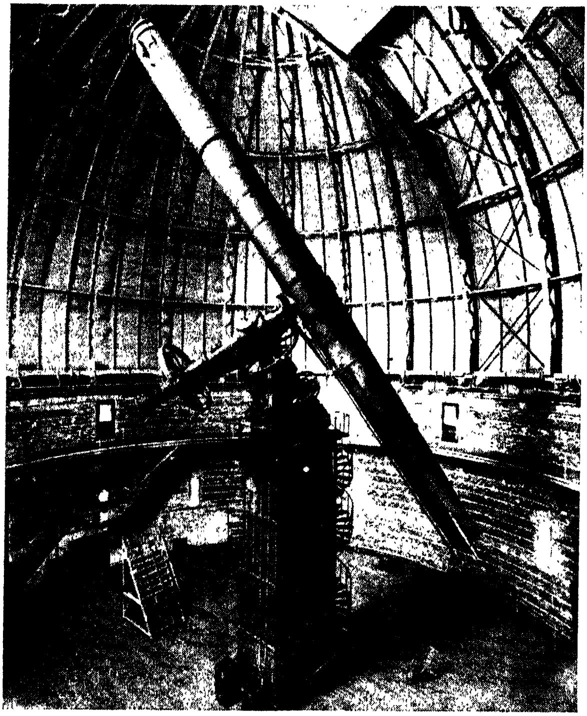
beam of light. An optical aid is necessary in order to gather more and to magnify any distant star or the sun. This is accomplished by the telescope. The essential feature of this instrument is a lens or mirror to intercept a large bundle of light rays and bend or reflect them to a point in order that they may be focused on the eye, a photographic plate, or scientific instruments.

Two essential types of telescope are now in use. The more common type is the one that employs a lens for focusing the light rays. It is known as the refracting telescope, since the rays are bent, or refracted, by passing through the curved lens. The same principle is employed in opera glasses or field glasses, with which everyone is familiar. The largest refracting telescope in use is at the Yerkes Observatory in Wisconsin. It has a front lens forty inches in diameter, or a total area of about 1,200 square inches. Assuming that the pupil of the eye has a maximum area of about 0.03 square inch, the telescope lens is approximately forty thousand times larger than the eye opening. Its light-gathering power is, therefore, that much greater than that of the unaided eye, and the light-gathering power is of utmost importance. We might say that this telescope is able to "see" an object forty thousand times too faint to be viewed with the naked eye.



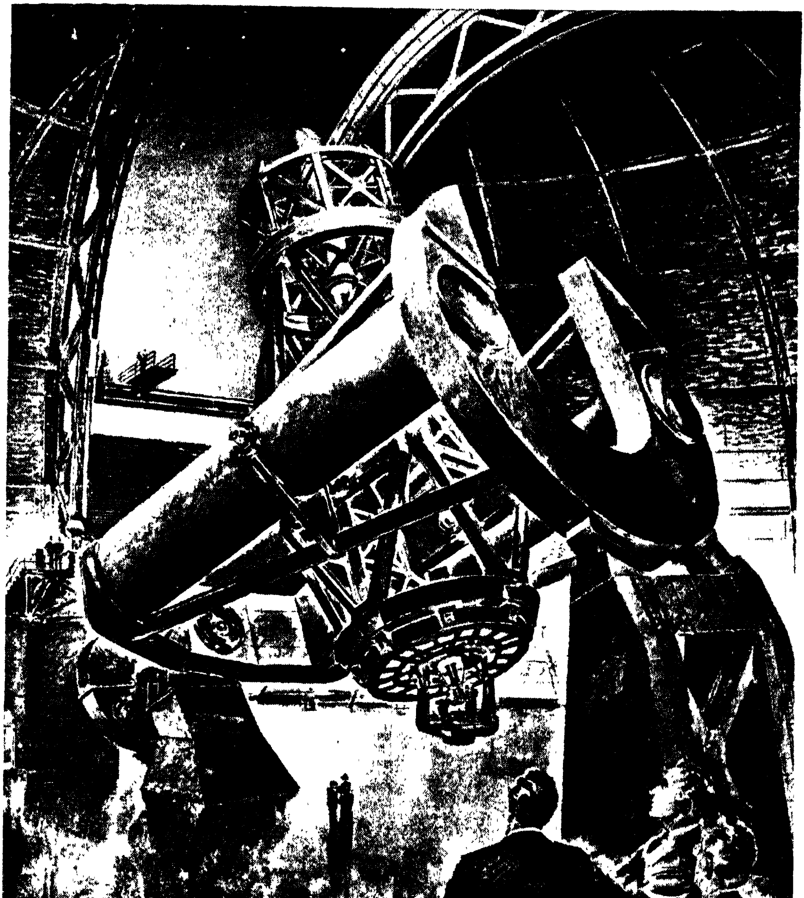
Illustrating the principles of a reflecting telescope above, and a refracting telescope below.

The magnifying power of a telescope is not of such significance as is its light-collecting property. The magnifying power



The large refracting telescope at the Yerkes Observatory with an objective lens 40 inches in diameter. (Science Service photograph.)

depends primarily upon the curvature of the great objective lens as compared to the curvature of the small lens used as the eyepiece. To be more specific, the magnification secured is equal to the focal length of the objective lens divided by the focal length of the eyepiece. The magnifying power of a telescope may be changed, then, by substituting eyepieces of different focal lengths, a relatively simple process. Just how much this magnification should be is determined to a large extent by the type of study and particular use of the telescope being made. There are certain definite limits, however, within which the magnification factors must come. For example, magnifications used by varying the eyepiece are most effective when they range between three times the diameter of the objective lens and forty times the diameter. In the case of the forty-inch Yerkes telescope, for example, a magnification of less than 3×40 does not make effective use of its great light-gathering power, and a magnifica-



The greatest telescope constructed by man, the 200-inch reflector at Mount Palomar, is shown in this artist's conception of the installation. The large mirror at the bottom intercepts the light and focuses it on a small mirror near the top of the framework which in turn reflects it back to the viewing mechanism. (Life Magazine photograph.)

tion of more than 40 × 40 would reveal no more detail of the object viewed than if the magnification were only 1,600.

The other type of telescope uses a curved mirror to gather the light and focus it to a point by reflection from its curved surface. After the light rays are focused to a point in front of the mirror, they are directed through an eyepiece at the side, or at some other position, by means of other mirrors set at the

proper angles. Such an instrument is called a reflecting telescope. The largest one in use today is at the Mount Wilson Observatory in California. It has a mirror 100 inches in diameter, which gives it a light-gathering power of about 260,000 times that of the unaided eye.

Of public interest as well as scientific importance is the new 200-inch reflecting telescope now nearing completion for the California Institute of Technology. Located at Mount Palomar, it is to date man's greatest telescope. It has a light-gathering power of about one million times that of the human eye. The mirror alone weighs twenty tons. It was constructed by the Corning Glass Works and required 372 days to cool and anneal after the glass had been heated for 40 days and poured into the forms. The entire telescope weighs about 1,500 tons, and a total expenditure of approximately $6,000,000 has been involved in its construction.

After the light has been gathered by the telescope, it must be analyzed and studied. Here the spectroscope has yielded most information. The spectroscope is an instrument that separates into its various colors or wave lengths the light passing through it, a process accomplished mainly by means of a transparent prism. Most people have observed the rainbow of colors produced when sunlight shines through a prism or a pane of cut glass. The rainbow in the sky is produced by raindrops refracting light into its different colors. Sunlight or any white light is made up of a large number of waves, each of different wave length. In passing through a prism the waves are bent, and those of the shorter lengths are bent more than those of the longer lengths. Thus the waves are separated. These separated waves are what produce the different colors.

When an electric spark is secured by causing a current to jump across a spark gap made by placing two pieces of a chemical element, such as iron, for example, in proximity to each other in a manner similar to a spark plug in a gasoline engine, a characteristic color of light for that element is always produced. A significant fact of nature is that each of the ninety-two chemical elements—gold, sodium, copper, oxygen, hydrogen, and so on—when brought to luminescence by an electric spark, will always give off a certain characteristic color or certain definite wave

lengths of light. Sodium, for example, will always emit a definite shade of yellow light. When this yellow light is passed through a spectroscope, it is separated into its various wave lengths. These wave lengths show up as bright lines in the spectrum, and no other light except that from activated sodium vapor will show these particular lines.

The characteristic wave lengths of light of an element are emitted when a sample of that element is activated in such manner as to cause a change of energy state within the atoms of that substance. This activation may be brought about in different ways, one of which is to produce an electric spark, as mentioned above. Another is to heat the element to a sufficiently high temperature. Each of the elements when heated to an incandescent gas will give a bright-line spectrum different from that of any other chemical element. In other words, the spectra constitute "identification tags" or "finger prints" of the different elements. Any light source may have its chemical composition determined by measuring accurately the lines that appear in its spectrum. Here, then, is a method of identifying the different chemical elements present in the stars by analyzing the light that those distant objects emit.

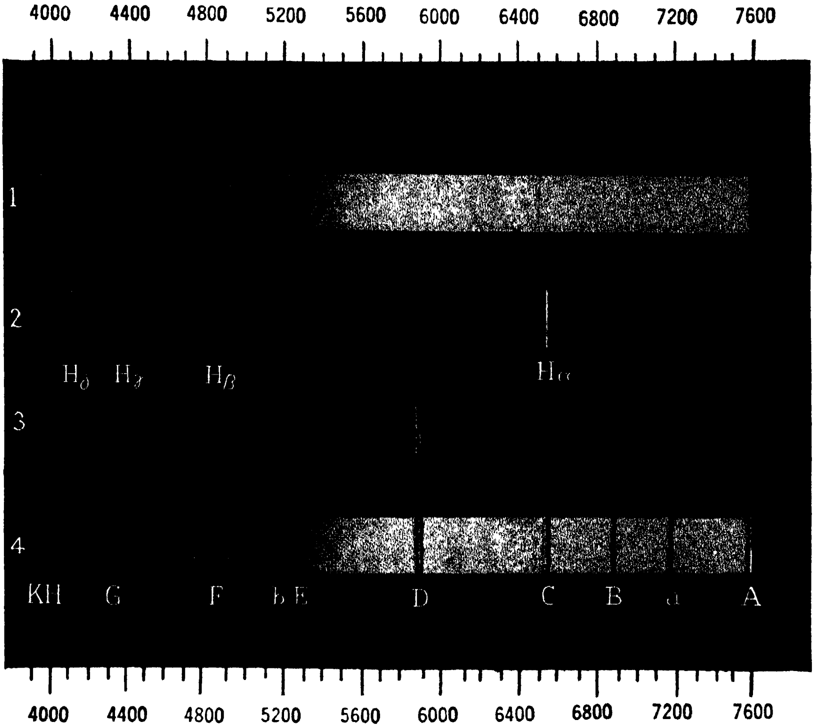
The elements have another significant property in their relation to light. People wear "sun glasses" to absorb some of the glare of sunlight from a bright pavement or the seashore. These glasses may be colored; frequently they are green. Such glasses have materials in them that absorb all the colors, or wave lengths, except green, which they transmit. More accurately stated, all the chemical elements may absorb light falling on them. The light that a particular element will absorb will be the exact color or have the same wave lengths that it would emit if it were heated to incandescence. For the sake of comparison with the bright lines mentioned above, let us consider again the element sodium. Should sodium in a gaseous state be illuminated with white light containing all wave lengths from another source, and the light then passed through a spectroscope, it would be seen that the spectrum is made up of a succession of colors of the white light with certain dark lines interspersed through it. The dark lines are caused by the absence of the wave lengths that have been absorbed by the element sodium, and they are found

to match exactly the bright lines produced in incandescent sodium vapor. This kind of spectrum is called a dark-line spectrum, and it is just as accurate in identifying an element as is the bright-line spectrum.

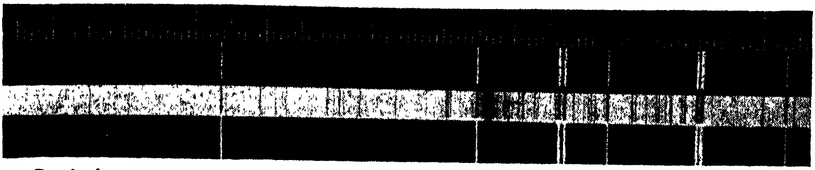
A more detailed spectrum of sunlight than is observed from an ordinary prism will show the continuous rainbow of colors with which we are all familiar; however, across this spectrum of color there will be a large number of dark lines. These are in reality gaps in the continuous spectrum. The dark lines in the sun's spectrum can be explained as follows: Light comes pouring out of the extremely hot photosphere of the sun. This light contains all colors, or wave lengths. It finally reaches the outer layers of the sun's surface, which are much cooler than the hot photosphere. Therefore, the gases making up these layers will absorb some of the wave lengths of light. Each element present will absorb its own wave lengths, and the rest of the light will pass on. By the time the light has run the gauntlet of all the elements in the cooler atmosphere of the sun and finally emerged into space, it is deficient in all wave lengths associated with those elements. These are the wave lengths that the elements absorb when cool and emit when hot.

These deficiencies produce the dark lines of the sun's spectrum; thus the lines are evidence of elements that constitute the outer portion of the sun. Furthermore, the lines that are missing from the light of the sun can almost always be matched with bright lines, or wave lengths, secured by heating in the laboratory known chemical elements on the earth. For example, the bright lines in the spectrum produced by heating the element sodium in a gas burner coincide exactly with some of the dark lines of the sun's spectrum. Sodium, therefore, must be present in the sun's surface to produce these dark lines. In this manner it is possible to determine the composition of the sun's surface layers by comparing the solar spectrum lines with the lines of the spectra of known elements on earth.

It is particularly significant that practically all the thousands of dark lines in the sun's spectrum are matched by bright lines obtained by heating the different elements that constitute the earth and all that is on it, such as hydrogen, helium, oxygen, nitrogen, iron, and copper. About sixty of the known elements on



A continuous spectrum of the colors produced by passing white light through a spectroscope is shown at the top. The second line shows the bright lines of the hydrogen, while the third shows two strong bright yellow lines in the sodium spectrum. The bottom line is the sun's spectrum showing a few of the dark or absorption lines properly lettered. Hydrogen and sodium are present in the outer surface of the sun since their bright lines match dark lines in the sun's spectrum. (Reprinted from Duncan's "Astronomy" by permission of Harper & Brothers.)



Bright-line spectrum of iron (above and below) compared with dark-line absorption spectrum of sun (center). The dark line in the sun's spectrum near the center of the photograph is the hydrogen B line shown at F in the color spectrum facing page 44. (Mount Wilson Observatory photograph.)

earth have been found in the sun. A still more remarkable fact is that the spectra of the different stars match closely that of the sun. This shows, of course, that the sun and stars are made up of practically the same materials and are in a comparable state of activation, or luminosity. The universe appears to be built of the same kinds of "bricks" throughout, whether its parts be the earth, sun, or remote stars.

Such uniformity in composition throughout the universe may be a revelation to many readers. Of the tens of thousands of stars that have been carefully examined with the spectroscope, only four different types according to composition have been found; and these four types blend into each other in orderly sequence. Evidently there is a close relationship among all stars, and the blending of types indicates that the only essential difference is one of stage of evolution, or point in the life cycle of the star.

Stellar Temperatures

Not only do starlight and sunlight reveal the kinds of matter in those heavenly bodies, but they also tell about the condition of that matter, *viz.*, its temperature and density. The temperatures of the surfaces of the sun and stars can be roughly inferred from the color of their light, like that of a hot poker or piece of steel in a furnace. If the color is red, the temperature will be about 2000°C. If it is orange, it will be about 4000°; yellow, about 6000°; white, about 10,000°; blue-white, 15,000 to 60,000°. With the exception of that represented by blue-white, all these temperatures are producible on the earth. From experiments on earth we can learn how color is related to temperature. Furthermore, the spectra of colors of different temperatures producible on the earth have been carefully analyzed, and such analysis

gives an accurate verification of the foregoing method for determining star temperatures.

Blue heat is observable in many of the stars. Most of the stars in the constellation Orion, for example, are blue-white. The star Sirius is at white heat; the sun is at yellow heat; Arcturus is orange; and the giant star Betelgeuse is red. All the other stars come within these temperature ranges. The temperature of the sun's surface is conservatively put at 6000°C., based upon the most accurate measurements. As a matter of comparison it may be noted that the temperature of molten iron inside a blast furnace is about 1900°C.; and temperatures as high as 3700°C. have been obtained in electric furnaces.

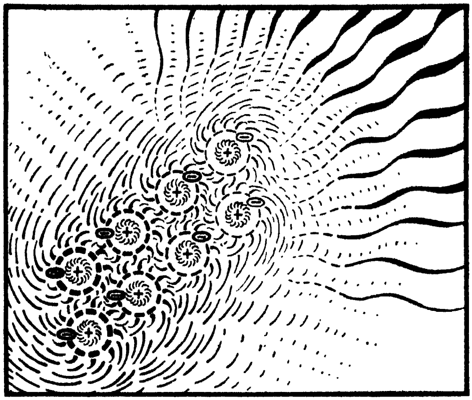
Heat from the Sun

When the temperature of the surface of the sun is known, it is possible to calculate the amount of heat energy that each square inch of surface radiates. A part of this energy reaches the earth; in fact all the heat that we use on the earth is received from the sun's radiation or has been so received in past ages. The amount of heat from the sun reaching the earth's surface is astounding. It is possible to measure and calculate it in several units, one of the most practical being the horsepower. The amount received is something like 130 trillion horsepower for the entire planet, an amount sufficient to allot to every individual on the earth a power supply equal to one-fifth the capacity of the great Muscle Shoals power plants!

Great as this figure is, it is absurdly small compared with the total energy radiated from the sun's surface. Some of this energy is absorbed by the upper layers of the earth's atmosphere and again radiated into space; also, only a small fraction of the sun's total radiation is received ninety-three million miles away by the relatively small earth, which, according to calculations, intercepts only about one part in two billions of the total amount.

We may wonder what the source of this enormous supply of energy is. Should you assume that it is simply a straightforward burning process by which the substance of the sun is being used up as fuel, you would be making no greater mistake than the early astronomers made for hundreds of years. However, a little calculation will show that if that assumption had been correct,

the sun could have lasted only about 1,500 years. Since the earth is at least two billion years old, and the sun much older, this, obviously, is not the answer. Another theory assumed that the sun is contracting and, by this process, producing its energy. This reasoning, first suggested by the German physicist Von Helmholtz in 1854, was based upon the well-known physical law that when bodies fall their initial potential energy is converted into heat energy. If the sun were contracting, all its individual parts would be falling toward the center with their potential energy being transformed into heat energy. If this is the manner in which the sun is generating its energy, it has sufficient volume to extend its life to about fifty million years. However, we know that the sun has been pouring forth heat and light for a much longer time than that, and all indications are that its potential future exceeds this figure. The sun's source of energy is some other condition, and Von Helmholtz's theory has been abandoned.



The source of the sun's energy is probably matter converted into radiant energy.

Although the exact internal changes by which the sun's energy is produced are not yet known, there is at present a theory that is satisfactory from the point of view of the time element and of all other observable facts. It is that matter is being annihilated and transformed into energy. We have seen on earth a few specialized instances of similar phenomena, *e.g.*, the atomic disintegration of radium. It is not improbable that the conversion of matter into energy goes on continuously in the sun's interior.

When matter is converted directly into energy, the amount liberated is enormous. Radium, for example, is continuously exploding, and a part of its substance being converted into energy. That a large amount of energy is liberated by even a

speck of radium is known to most people. If one gram of coal were to be completely changed into energy, it would furnish sufficient power to drive H.M.S. *Queen Elizabeth* across the Atlantic. This idea of the conversion of matter into energy is in accord with Einstein's conception that the mass of an element may be completely transformed into energy. Such changes are believed to be taking place continuously in the sun.

However, even this explanation of the supply house for the sun's enormous radiation requires that some of its mass be used up. According to Einstein, one gram of matter, regardless of its composition, when entirely converted into energy will produce 2.15×10^{13} calories of heat. On this basis, in order for the sun to maintain its liberal expenditure of energy, it must be losing mass at the rate of four million tons every second. Thus it is decreasing in mass one-thousandth of one per cent in 150 million years. This explanation of the sun's source of energy would allow it sufficient length of life to conform to what its age is believed to be. It does indicate, however, that the sun is slowly decreasing in size and that eons of time hence it will join the ranks of the darkened dwarf stars. No more light and heat will then radiate from it, and the earth will become a cold, dead world.

Even with this enormous brightness of the sun, many of the stars have an intrinsic brightness that is greater. The star with the greatest brightness of all, with the exception of occasional supernovae, is one called S Doradus, visible in the Southern Hemisphere. S Doradus gives out at least three hundred thousand times as much radiation as does the sun. If it were suddenly to replace the sun, everything on the earth would be roasted in a few seconds and soon turned into vapor—sea, rocks, earth, and all. The bright star Sirius is twenty-seven times as luminous as the sun. In fact, most of the stars that can be seen with the unaided eye are glowing masses, intrinsically brighter than our sun.

But many of the stars making up the universe are much less luminous than the sun is. The nearest of all known stars, Proxima, gives not only about one-twenty thousandth part as much light as the sun does. If it were to replace the sun, the earth would become exceedingly cold. In a short time there would be mountains of ice where we now have liquid water, and

any rivers existing would run with air that had cooled and condensed to a liquid as a result of the low temperature.

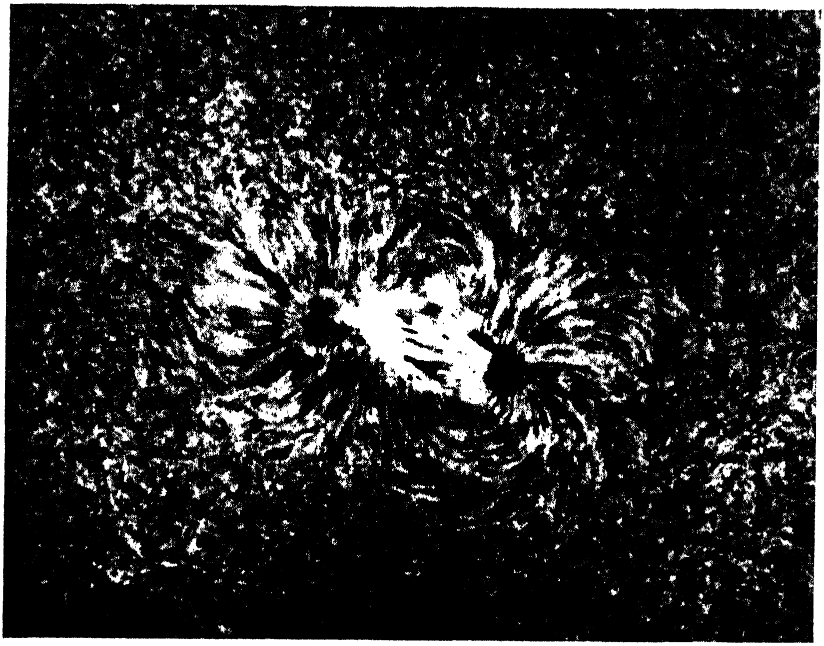
The Sun Is an Incandescent Gas

From the foregoing statements it has probably been surmised that the sun is a huge mass of flaming gas. This is now known to be true—in startling contrast to the solidity that we associate with the earth.

The sun's gaseous condition is indicated by a number of observable phenomena, one of which is that the sunspots do not move with uniform motion as the sun rotates. We all know that the relatively rigid surface of the earth carries all points on it around uniformly, so that a complete rotation at Chicago occurs in exactly the same time as a complete rotation at the equator. However, the sun's equator rotates much faster than do its northern and southern latitudes, as indicated by the speed of rotation of sunspots as identification points on its surface. Obviously, this can mean only that the sun is not a solid body.

Neither is it a liquid; the temperatures are too great. We have previously noted the temperature of the sun's surface; its interior is even hotter. Astronomers have no way of determining directly what these interior temperatures are, but computations involving the sun's weight and interior pressures put the maximum at about 40,000,000°C. This is so much more intense than any temperatures producible on the earth that we have no way of comprehending it. Obviously any material, even platinum or the diamond, would be instantly converted into a superheated gas at such heat. At the temperatures found in the sun, substances exist only as gases.

In addition, a direct examination of the sun's surface reveals it to be flaming gas. When looked at normally the sun appears to be a rather quiescent body; however, should we approach it closely and not be thoroughly roasted in the process, we could behold its seething character. The seething condition is revealed to us at the much safer distance here on earth when the sun is totally eclipsed by the moon. The sun's main disk is blocked out, and its surface shown up by contrast against the dark moon. We can then see the surface, ordinarily obliterated by the intense glare of the sun proper.



Closeup of sun's surface and a bipolar sunspot of Aug. 30, 1924. Sunspots consist of huge cyclones of vaporous masses of hydrogen, helium, magnesium, iron, and other elements, in which the gases flow inward near the base and outward at higher levels. All sunspots are accompanied by a strong magnetic field, and when appearing in pairs, as shown above, they usually have an opposite magnetic polarity. Dark centers are about the diameter of the earth. (Mount Wilson Observatory photograph.)

Terrific vapor storms, not unlike the flames of a forest fire, sweep over the sun's surface. Great jets of lighter incandescent gases are continuously leaping far out from the sun. Gases of the heavier metals also shoot out but not to such great distances. These great jets are known as prominences, and they are easily visible at any total eclipse. Frequently they ascend to hundreds of thousands of miles. Some prominences resemble rockets and jets of fire which travel at speeds up to 500,000 miles per hour. Others are more quiescent, appearing on the sun's surface in the shape of massive pillars or pyramids. The photograph at the beginning of the chapter shows a closeup of some of these prominences.

The prominences seem to rise from a layer of lighter gases, mostly hydrogen and helium. They constitute a layer known as the chromosphere. This term means the "color" sphere, for it is

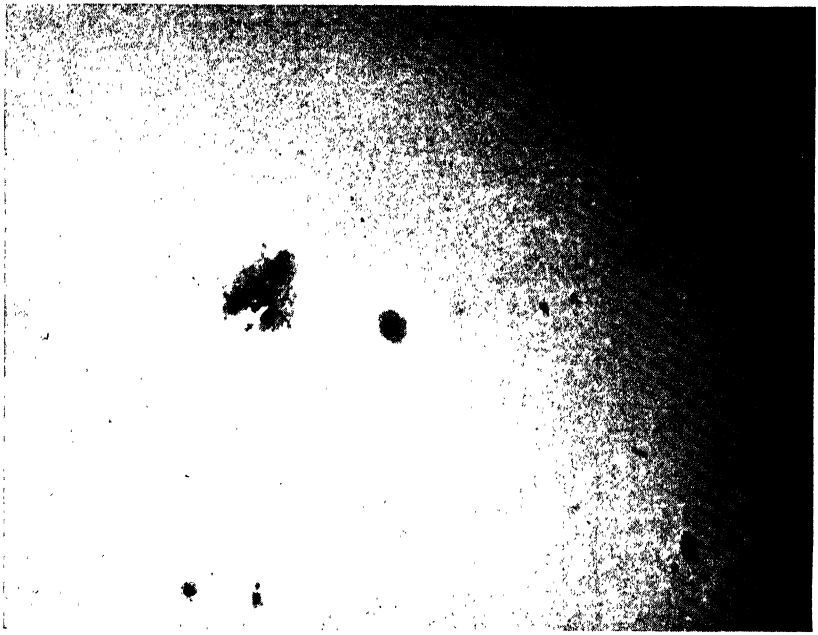
here that we find a region of brilliant red color, which covers the sun's surface for a thickness of several thousand miles. Beneath the chromosphere is the true surface of the sun, the photosphere, or "light" sphere. This is the part of the sun that we see normally, the part that contains the sunspots. Extending out beyond the chromosphere and far beyond the prominences is a third layer consisting of soft, pearly light, the corona. It surrounds the sun completely to depths of 300,000 to 500,000 miles and is visible to the unaided eye only at times of total eclipse. The corona may be produced by sunlight reflected from a sort of cosmic dust surrounding the sun.

Sunspots

Frequently enormous agitation is seen in the photosphere. These disturbances often cover millions of square miles in area and constitute what are called sunspots, which usually last for two or three days but sometimes continue for several weeks and occasionally for months. Sunspots are never seen near the poles but always within forty-five degrees of the sun's equator. They frequently move across the sun's surface and also rotate with it. They are usually thought of as "dark," as that is the way that they appear; however, it is only by comparison with the greater brightness of the rest of the sun that they look dark. Actually they are quite bright, only a little less so than the rest of the sun. At times they appear to be great depressions in the sun's surface.

Sunspots usually begin in the higher latitudes and move toward the sun's equator where they disappear. Their number varies greatly from time to time. One period of variation has been found of about eleven years' duration. At the climax of this eleven-year cycle, the sun is never free of spots, and sometimes a hundred or more are visible at once. At its lowest ebb, weeks, even months, may pass without a single spot. Fifteen of these eleven-year periods of variation have now been recorded, and the cycle is well established.

Just what causes sunspots is not definitely known; however, it is probably something connected with the condition of the sun itself rather than any outside influence. Sunspots are centers of electromagnetic fields. They often occur in pairs, and a strong magnetic field surrounds them, one of the spots being a north



Sunspot which caused the magnetic disturbance on Mar. 24, 1940. The earth is about the diameter of the dark center in the nearly round near-by spot. (Science Service photograph.)

and the other a south magnetic pole. This condition may have something to do with the effects that they have on the earth.

Sunspots affect terrestrial magnetism. The earth's magnetic field is never quite constant; but the most pronounced disturbances are the magnetic storms. At these times, the direction in which the compass points has been known to change as much as three degrees in as many minutes. Magnetic storms occur at times of sunspot maximum. At the same time an interference with radio reception has also been observed. In the early part of 1940 occurred one of the worst magnetic storms known to modern man. It caused a disruption of telegraph service, disturbed radio communications enormously, and interfered with long-distance telephony. A particularly large sunspot group was visible at the time, and it evidently had a specific connection.

There seems to be a decided relation also between sunspot variation and weather cycles, for it is fairly well established that there is an eleven-year weather cycle, which coincides with

the sunspot period. This has been shown by comparing the Weather Bureau records with sunspot activity. Periods of sunspot maximum are accompanied by somewhat cooler weather and greater average annual rainfall. With the increase and decrease of numbers of sunspots, summer climates gradually change from cool and wet to hot and dry and then back again. Likewise, winter climates vary somewhat accordingly in their coldness.

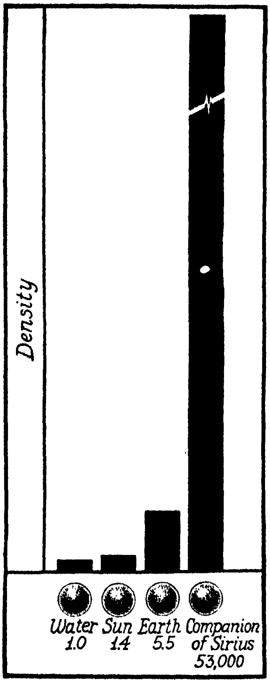

An aurora borealis display.

Northern lights, or aurora borealis, displays are also much more marked during times of greatest number of sunspots. Streams of electrons, negative electric charges, which pour out from the sun at these times, after wandering around for a few days, reach the upper layers of the earth's atmosphere and intensify the electric discharge, causing the gases of the upper atmosphere to glow in a manner not unlike the functioning of a neon sign. The glowing gases produce the beautiful and constantly changing display of lights seen in the northern sky.

Densities

How heavy is this material that makes up the sun? We need no longer guess at the answer, as astronomers have been able to measure its weight with a high degree of accuracy. For purposes of comparison of one object with another, density is a more convenient and more accurate measurement than is total weight. Density means the weight of a unit volume of a substance. For example, a pint of water weighs about one pound. To be more specific, a cubic centimeter of water at a temperature of 4°C. weighs one gram, the cubic centimeter being one unit of volume and the gram being the unit of weight in the metric system of measurement. Therefore, the density of water is 1. On the other hand, one cubic centimeter of steel weighs 7.8 grams; accord-

ingly, the density of steel is 7.8. The density of the entire earth has been measured; it is 5.5; that is, the earth is 5.5 times as heavy as it would be if it were composed entirely of water.



Comparison of densities.

The density of the sun has been measured by two distinct and different methods. One, the simpler, involves calculation of the sun's total mass. This is accomplished by d e t e r m i n i n g the sun's gravitational attraction for the earth, which keeps the earth in its orbit while it is revolving around the sun. With the sun's gravitational force as a constant, the sun's mass is calculated. It is then simple arithmetic to determine its density, or weight per *average* unit volume, by dividing the sun's total volume (which has been accurately measured) into its mass. The sun's average density thus obtained is 1.4, or about one-fourth the density of the earth.

The other method makes use of a unique consideration. It is, as may be surmised, a study of the light emitted from the sun. The lines of the spectrum have their degree of fundamental sharpness determined by the density of the material emitting the light even when the spectroscope is in perfect focus. The sharper or better defined these lines are in the spectrum, the less dense is the luminous source; the fuzzier the lines, the greater is the density. This holds true for light coming from the stars as well as for that from the sun. Studies of this type have revealed that the sun's density ranges between quite wide limits, depending upon the part of the sun from which the particular light comes. The extreme outer portions have a very low density, comparable with that of a rarefied gas; the interior portions are exceedingly heavy with densities greater than that of any substance known on the earth. The bright star Sirius has an average density of 0.75, and its less bright companion star has a density of 53,000. A tablespoonful of this material at the earth's surface would

weigh a ton. The great red star Betelgeuse is only about one-millionth as dense as water, which makes it a glowing, rarefied gas, similar in density to the contents of luminous neon tubes.

The largest of the stars are those of lowest density, while the smaller ones usually have the greatest. When average density is multiplied by volume, the star's total weight or mass is obtained. Although a great variation exists in the densities of different stars, there is remarkable uniformity among them in total weight. If the sun's weight is taken as a standard, most of the other stars will be relatively close to this weight. The most massive will be not more than about four hundred times as heavy as the sun, and the lightest will be not less than approximately one-four-hundredths of the sun's weight. Nature seems to control the mass in a star; if the mass is too great, the star perhaps breaks up; if it is too small, the star does not radiate enough heat and light to be visible as such.

Multiple Stars and Variation

A general view of the heavens does not reveal that many of the stars have companion stars relatively near them; closer observation, however, shows such to be the case. About half the stars are twins; a considerable number are triplets; and a few are quadruplets. (The sun, however, is a lone star, having no near companion.) More than a hundred years ago Sir William Herschel concluded that some of the stars that appeared as single points of light were really double. Even the telescopes of his day would resolve a few of them into two points of light. When the spectroscope began later to be used with larger telescopes, it was found that many other stars gave two spectra, one closely superimposed upon the other, indicating that the source was a double star. These spectra often show that the double stars are revolving about each other in elliptical paths. The most remarkable discovery of this sort concerned the bright star Sirius. It was found to follow a wavy-line path as it moved through space. This could be explained only by the assumption that it possessed a twin star and that the two were revolving about a common center of gravity. The common center of gravity moves in a smooth curve, but each revolving star follows a wavy path

as the stars revolve about each other. When more powerful telescopes were built, Sirius's faint companion star was discovered in 1862.

Some stars exhibit peculiar variations in brightness because of an eclipse effect. The eclipsing type of star has a fainter twin star, and the two revolve about each other. The fainter one comes between the bright one and our earth, producing a partial eclipse of the brighter, visible star. This makes the star less bright than it is at other times. About 200 pairs of the eclipsing type are known.

Another remarkable property is common to some stars, this being that they vary in their intrinsic brightness. It was once believed that the stars were immutable and in every way eternal; but some 5 per cent have in recent years been discovered to vary in actual candle power. This is not caused by the eclipsing mentioned above but must be due to action within the star itself.

The most obvious intrinsic variations are seen in novae, which are "new" or temporary stars that have flared up. Some novae may be entirely invisible, then rather suddenly become relatively bright. Their brightness usually increases rapidly to a maximum within a few days, after which it slowly dies down. When this increasing brightness begins, the star starts to expand enormously; the glowing mass is spread out; and its candle power increases greatly, sometimes 300,000 or more times than it was at the beginning. As the expansion increases, the gas becomes too rarefied to shine powerfully, and the star begins to decline in candle power. Why stars should expand and increase their brightness this way no one knows, except that some sort of "explosion" takes place in which the outburst of energy exceeds all other known physical catastrophes.

In addition to the novae there are other kinds of variable stars. The most important are those which pulsate at more or less regular intervals. Some of these regular intervals are about 400 days in length; others are as short as a few hours. One star, visible in the north latitudes and called Mira (The Wonderful), is sometimes almost as bright as the North Star. This great brilliance lasts for about 15 days; then it decreases for about 90 days, when it becomes invisible to the unaided eye; it so

remains for about 200 days; then in about 90 days more it increases again to its original brightness.

The variable stars of very short periods are known as the Cepheids. They are stars of exceedingly great candle power which pulsate in brightness every few hours or a few days at most. A remarkable relationship has been discovered between actual candle power of such stars and the exact period of their variation. Thus the intrinsic brightness of a Cepheid may be quickly and accurately determined by measuring its period of variation. By knowing the brightness of a star, astronomers are able to calculate its distance from the earth. Therefore, the distances of most of the Cepheids have been determined by measuring their periods of variation. This system has been of inestimable value in determining the distances from the earth of a great many star groups that contain Cepheids, as calculation that could not have been accurately made by other known methods of measuring star distances.

Numbers and Sizes

Contrary to popular thought, the number of stars in the heavens is by no means infinite, although it is very great. If one were of a mind to, he could count in the northern sky on a clear, moonless night about 3,000 stars visible to the naked eye. About the same number appear in the sky above the Southern Hemisphere, making a total of some 6,000 stars visible to the eye. With the aid of a good pair of field glasses the number is increased to about 100,000. The 100-inch telescope at Mount Wilson has revealed by means of photography about one billion stars. Dr. C. G. Abbot, noted astronomer of the Smithsonian Institution, has estimated that there must be at least 30 to 40 billions; Sir James Jeans claims that the number is near 100 to 200 billions!

The stars vary immensely in size. Our sun might be thought of as an average-size star, having a diameter of approximately 860,000 miles. This is a rather sizable object—so large, in fact, that, were it hollowed out, a million cubical blocks each the volume of the earth could be placed inside with still room to spare. Yet many of the stars are much larger than the sun. One of the largest is Antares, with a diameter 480 times that of the

sun. The space that it occupies could hold over fifty million of our suns. Another star of great magnitude is Betelgeuse, which is evidently a pulsating star with a diameter varying between 320 and 540 times that of the sun.

By far the largest star known to man is one discovered in 1938 by astronomers at the Yerkes Observatory. It is the "ghost" companion of Epsilon in the constellation Auriga. It has never been seen and probably never will be seen by man. Its temperature is so low that it does not shine with visible light; however, it does emit infrared radiation, or heat waves, similar to the heat waves emitted by a heated household iron. This invisible radiation has been detected with the thermocouple and also by special infrared photographs which are sensitive to radiant energy. A very minute and careful study of these radiations has revealed that the star is unbelievably large—about three thousand times as large as the sun. The solar system, including the sun and the planets as far out as Uranus, could fit within the space of this infrared star.

Other stars vary in size from these largest to those which are much smaller than the sun. The small ones are called "dwarfs." The smallest of all known stars are only about the size of the earth, or about 8,000 miles in diameter.

The sizes of the stars are not mere random quantities but are closely related to their physical states. The largest are the low-temperature, tenuous, red stars. As the sizes get progressively smaller, the stars get hotter, until a diameter of about ten times the sun is reached, this being the case of the blue-white stars. Then as the sizes continue to decrease, the temperatures begin to drop. The sun, for example, is much cooler than the blue-white stars and is classified on the color-temperature chart as a yellow star. Stars smaller than the sun become cooler and redder. These are the well-known dwarfs. However, it seems that the very smallest of the dwarfs are again very hot and give off blue-white light.

The clear indication of this relationship between size and temperature is that stars have definite life cycles. Now, we may make brief note of some of the ideas about star evolution and life cycles. These ideas are only theoretical, but they are deductions based upon thoroughly verified observations. The young

stars are believed to be the great red masses, like Betelgeuse, formed probably out of nebulous matter in space. As times goes on, they continue to contract, and the temperature rises. Finally the temperature reaches a point where the star shines with a brilliant white-hot shimmer, and it may now be considered a full-grown adult. Such a star is Vega, with many thousand times the intrinsic brightness of the sun. Stars like Vega are not only objects of beauty but are internally the most active of all cosmic bodies.

As the star continues to radiate energy at such an enormous rate, it begins to cool down and to shrink to a smaller size. The color becomes yellowish, and the star passes into old age. Such a star is the sun. After more time, a star gets still smaller, and the continued radiation of energy makes it colder and redder. These are the old-age dwarfs. Before final extinction the continued contraction of the star causes it once again to raise its temperature and shine once more with white brilliance, this probably being the final flare of its celestial existence as a star. Doubtless a time interval of hundreds of billions of years is required for this life cycle.

Star Distances

One of the most astonishing facts about the stars is their great distances from the earth. The nearest star to the earth other than our sun is Proxima, out at a distance of 4.3 light-years —about a million times as far away as the nearest planet. The light-year is the distance, it will be recalled, that light travels in 1 year's time, going at the velocity of 186,284 miles per second. In 4.3 years' time it would travel about 25,000,000,000,000 miles, which is the distance to the nearest star. Sirius, apparently the brightest star in the sky and visible in northern latitudes, at certain times of the year is 8.7 light-years away from the earth.

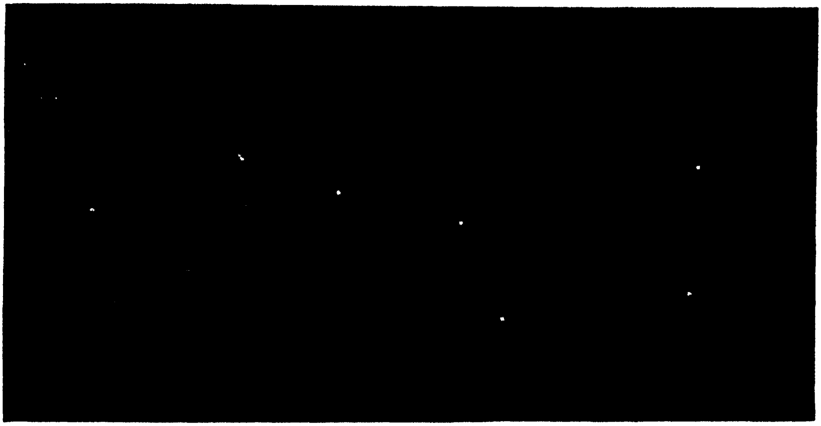
Few stars are less than 10 light-years distant. The Pole Star is about 1,000 light-years away, a distance subject to measurement, however, only with a considerable margin of error. Arcturus, the star whose light was used to open the Century of Progress Exposition in Chicago in 1933, is approximately 40 light-years away. The light that fell on the photocell to open the Exposition gates in Chicago that night started on its long

journey to the earth in 1893 when the first Chicago World's Fair was held; and during the intervening years it had been hurrying through space at the enormous speed noted above until it was intercepted by the tiny electric eye, or photoelectric cell, at the time of the auspicious occasion just mentioned. Other stars may be approximately 50,000 light-years away from the earth. These distances are enormous, yet they include only the stars that belong to our own galaxy.

It must be evident to everyone that the heavens are immense, and that stellar space is sparsely populated. The great distances and enormous sizes are really beyond comprehension. Perhaps a simple comparison will make them a little more understandable. Imagine the stars to be reduced until the average star, say the sun, is the size of an orange, and all distances diminished accordingly. Then, to represent a true picture of the distribution of the stars, one orange representing the sun could be placed in New York; another representing a star would need to be placed in New Orleans; one for a second star in Los Angeles; and another for a third star in Alaska. The orange to represent the farthermost star on this scale of reduction would need to be placed out beyond where the sun now is. Should our Galaxy of stars be so reduced as to allow us to comprehend its component parts, it still would remain enormous.

Motions and Positions

One other condition of the stars seems quite astonishing when their usual appearance is considered: They are all in motion. We have been taught to call stars "fixed," as contrasted to the "moving" stars, or planets. The stars making up the Big Dipper, for example, are in the same relative position to each other now that they were when Columbus discovered America or even when early Greek civilization was at its height. But the fixed stars do move; and in this respect things are not what they seem. The explanation of this apparent paradox is that the stars are so far away that within the lifetime of any man or even within several centuries their motion is not perceptible to the unaided eye. But after thousands of years the Big Dipper will no longer look as it does now, for the stars will have shifted their positions. Time eventually shows clearly that the stars are

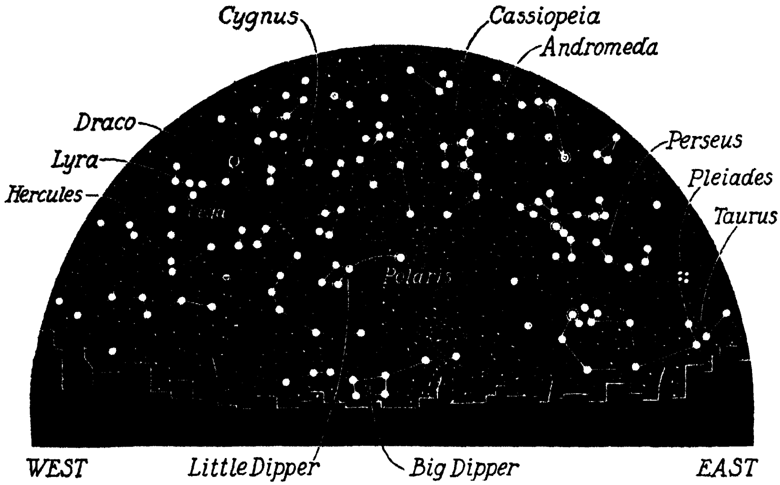


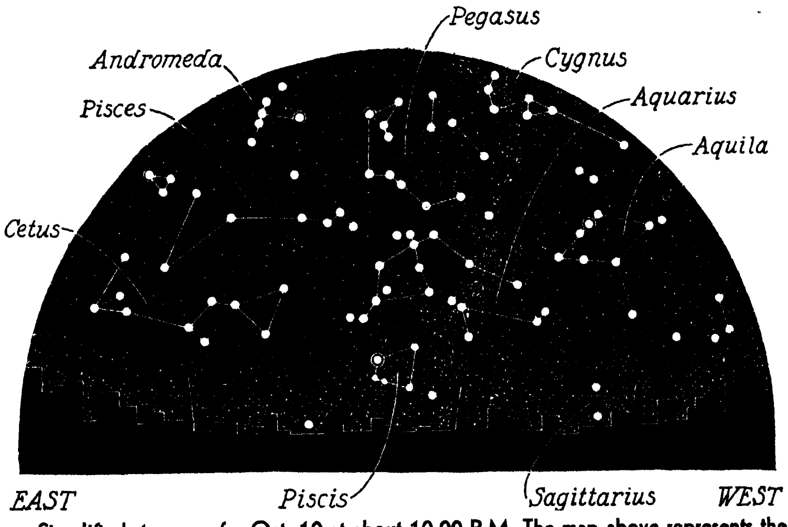
The Big Dipper as the eye sees it. In this special photograph all stars not in the major constellation have been retouched out. (Science Service photograph.)

moving. About fifty stars are known to be moving with respect to the solar system at velocities exceeding fifty miles per second; but a majority move at the rate of less than twenty-five miles per second relative to the solar system. Some stars are moving in a general direction toward the solar system; some, away from our sun; still others are moving across the line of sight. The sun itself is hurtling through space, carrying the planets along with it. It is moving at a speed of about twelve miles per second toward an apex in the sky that is near the constellation Hercules.

By referring again to the model of the oranges, it is possible to get a general notion of the magnitude and significance of these star motions. If velocities are reduced in the same ratio as stars being represented by oranges, then a rough approximation of stellar velocities would be that the oranges move about one foot per century. It can be readily understood that if a person in New York were able to see the orange in Los Angeles with the unaided eye, he would not be able to see any motion of it even within a lifetime. However, were he able to observe it with one of our best telescopes, he could probably detect the movement within ten to twenty years. Perhaps this comparison may also make clear the reason why the stars seldom collide with each other in their movements through space.

As far as we are concerned, then, the stars retain their fixed positions during a human lifetime. Anyone with a simple star





Simplified star maps for Oct. 10 at about 10:00 P.M. The map above represents the half of the sky from a point overhead to the northern horizon and the lower one the sky from the zenith to the southern horizon.

map may determine the names and respective positions of some of the brightest ones. The two simplified maps shown are for use in northern latitudes, and each represents about half the celestial dome from a point directly overhead to the northern and southern horizons, respectively. Both maps are to be used at about the hour indicated and within a few days of the date. To

use them, face north or south (depending upon the map selected) at the appointed time, and hold the map horizontally with the northern horizon pointing to the north or the southern pointing south. The stars in the sky overhead will be in the relative positions indicated on the map below in your hands.

Even such simplified star maps as the ones shown, and others of a more detailed and specific type which show the sky at different times of the year, may assist the amateur stargazer in becoming familiar with a few of the brighter stars and constellations as they appear in the sky. Instead of the simple maps represented here, astronomers use a more exact system of right ascension and declination lines across the sky. These are imaginary lines from known reference points which correspond respectively to north-south longitude lines and parallels of latitude on the earth. The stars' positions are specified in this system by so many hours of right ascension and so many degrees north or south of the celestial equator. When these readings for a specific star are set off on a telescope that has been properly calibrated and the readings centered, the telescope points directly at the star.

The Galaxy

All the stars that have been mentioned above belong to a large group, or system, that includes the sun and is referred to as the Galaxy. The Galaxy is of such immensity that all the stars visible to the unaided eye, with the exception of a few nebulae, belong to it. Even though it occupies an enormous space in the universe, its form and shape are well recognized.

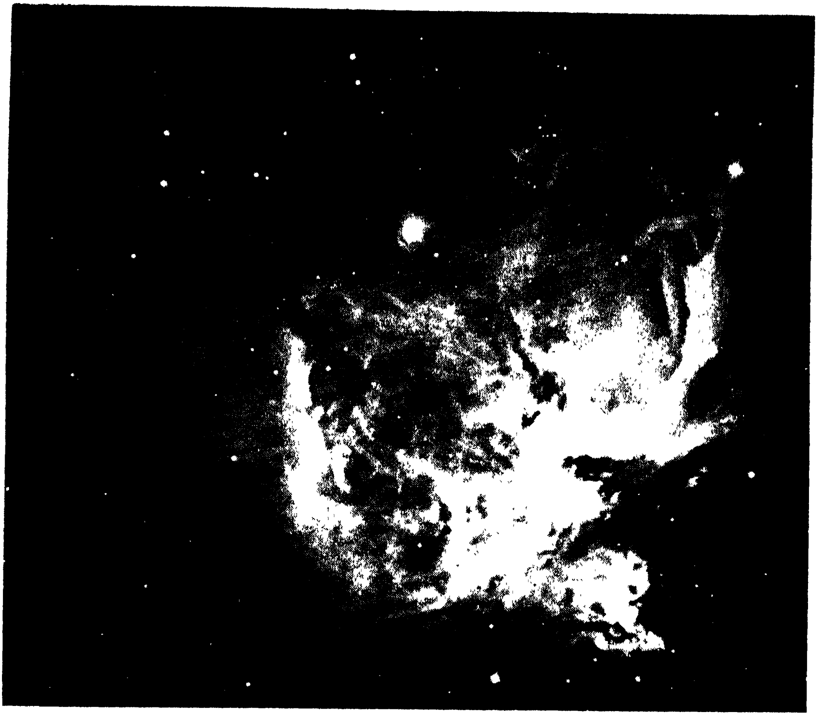
The form of the Galaxy may be understood by anyone who looks at the sky and notices the distribution of the visible stars. Ancient peoples as far back as recorded history had observed a misty, cloudlike path extending across the heavens like the arc of a great circle. This fleecy and beautiful path has figured in the legends of primitive man and has enabled astronomers to determine the shape of the Galaxy. The North American Indians call it the "Trail of the Spirits," believing it to be the path of the dead to the Happy Hunting Grounds. We call it the "Milky Way." Although it looks to the naked eye like a hazy band of light, telescopes show that it is made up of more stars than are found in other parts of the sky. Furthermore, in any

part of the sky extending from either side of the Milky Way the number of stars decreases as the distance from the misty band increases. A study of this grouping of the stars reveals to us the shape of the Galaxy.

When the positions of all the stars are mapped out, it is found that the Galaxy is shaped somewhat like a disk or watch or pinwheel, bulging a little near the center like a double convex lens. This pinwheel disk is approximately 80,000 light-years in diameter, and the distance from "front" to "back" is about 10,000 light-years. The sun occupies a position in this system about 30,000 light-years away from the center of the larger, or rotation, diameter. As we look out along the radius of the disk, we see more stars, not because they are closer together but because of the greater distance involved. They appear as a bright band, the Milky Way. As we look out along the axis of the disk, we see fewer stars because of the relatively shorter distances, and the sky appears to be less thickly populated.

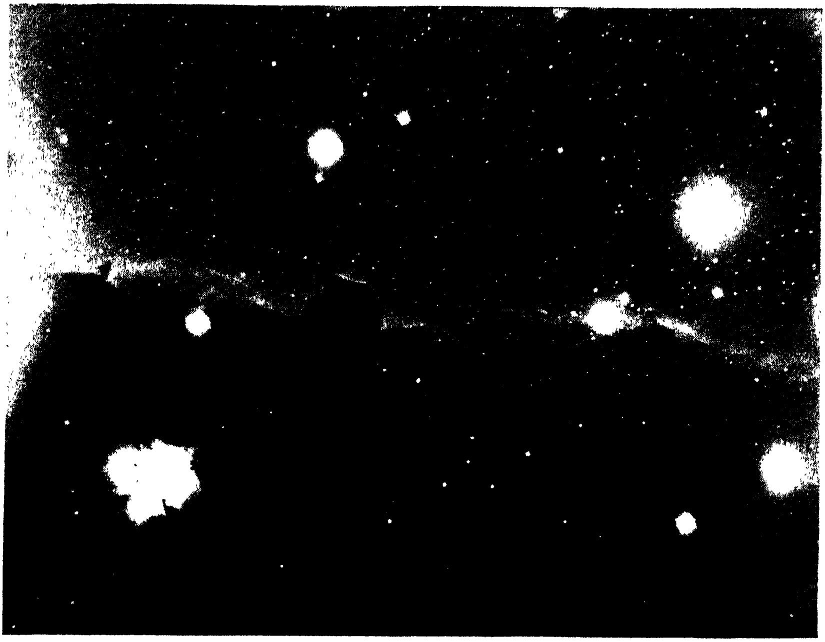
The motion of the stars in the Galaxy is not all haphazard and at random. In addition to the individual motions of a great many of the stars, as previously noted, it now seems to be well established that the whole galactic system is rotating around a center, much as a pinwheel rotates around its hub. This rotation of the great wheel of stars whirls the sun through space at a rate of about 200 miles per second. However, the wheel is so vast that the sun must travel at this rate for about 220 million years before it makes one complete revolution. But the ages of the stars are such that the celestial wheel must already have made many thousands of complete turns!

In addition to the multitude of single and double stars the Galaxy includes many star clusters. These are star groups consisting of a few hundred and, in some cases, a few thousand stars. Probably the most popularly known and most interesting of such clusters are the Pleiades, several hundred closely grouped stars. This cluster is prominent in the eastern and southern skies of the Northern Hemisphere during the month of November. About seven of them are visible to the unaided eye. When photographed through the telescope these brighter ones are seen to be surrounded by a radiant nebulosity.



Bright Nebula in Orion. (Mount Wilson Observatory photograph.)

Present also in the Galaxy are many bright and dark nebulae. They seem to be huge, apparently gaseous clouds. Their densities are exceedingly small, probably in the order of about one-millionth that of the best vacuum that can be produced on the earth. The bright nebulae are believed to be illuminated by bright stars relatively near them, and their cosmic dust particles as well as their gases reflect the starlight so as to make them visible. Such is the great gaseous nebula in Orion. The dark nebulae are probably similar in general composition to the bright ones except that there are no bright stars near by to illuminate them and also that their dust particles are of such size as to absorb the light rather than to reflect or transmit it, as a dust or smoke screen over a community on the earth shuts out the sunlight. One of the most spectacular of these dark nebulae is the Horse's-head in the constellation Orion.

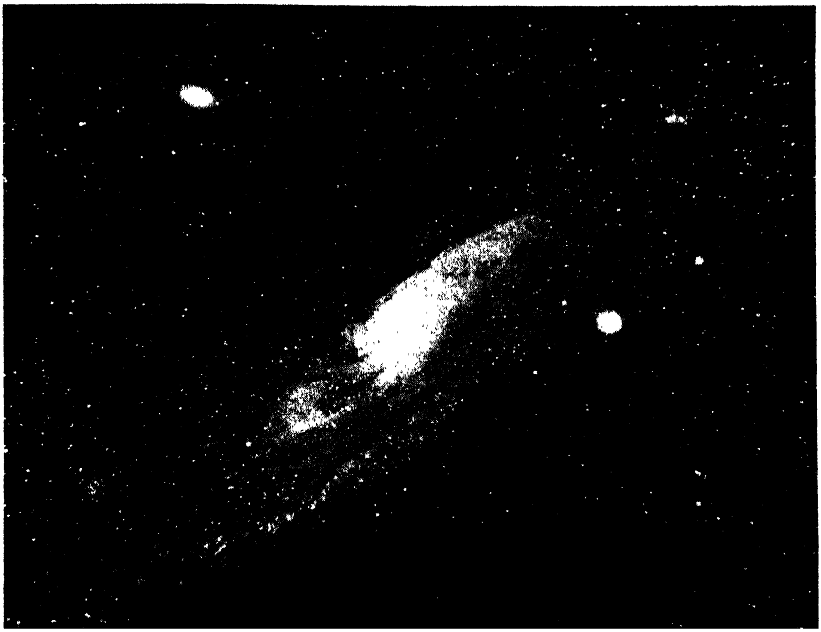


Horse's-head Nebula in Orion. (Mount Wilson Observatory.)

Island Universes

In addition to the system of stars that constitute the Galaxy there are still other systems. Now we must be able to stretch our imaginations to the cracking point, for far out beyond the limits of our Galaxy other groups, or nebulae, of amazing import have been discovered. The brightest and best known of these is the Great Nebula in Andromeda. It is at least 700,000 light-years away from our Galaxy. The large telescopes reveal it to be made up of millions of separate stars. Recent measurements with the photoelectric cell indicate that it is probably as large as our own Galaxy and possibly contains about as many stars, which is in the order of billions. It forms what is usually called an island universe, evidently not greatly unlike our Galaxy in size and shape.

Photographs of other parts of the sky, using the great telescopes, show that there are other distant star groups, other island universes. Several millions of them are known to astronomers, and they have been found scattered throughout all parts



The Great Spiral Nebula in Andromeda.

of the sky. The stars that compose some of these groups are known, from a study of their spectra, to be revolving, like giant pinwheels, around a center. The groups themselves are usually called spiral nebulae, although the term nebulae is unfortunate, since it originally meant nebulous clouds in our own Galaxy. Some of these groups present flat sides to the earth; others present disk edges; and still others show the planes of their rotation axes at various angles. Most of them appear to be disklike in shape and spiraling in a given direction.

Just how many of these island universes exist is unknown, but probably there are many millions. The most distant of all known to man is one that has been photographed with the Mount Wilson telescope. It is 500 million light-years from the earth! What lies beyond this no one at present knows. We have no reason to believe, simply because it is not possible for man to see greater distances, that this is the end of space. Astronomers look forward to the completion of the new 200-inch telescope with which they will be able to reach farther out into stellar space. In a few years those who follow the developments of

science will probably read many interesting stories of what has been learned about the heavens far beyond the limits of our present knowledge.

REFERENCES FOR MORE EXTENDED READING

JEANS, SIR JAMES: "Through Time and Space," The Macmillan Company, New York, 1934.

Included in this book is a series of popular lectures delivered before the Royal Institution by one of England's most noted astronomers. Chapters VI, VII, VIII contain a wealth of information imparted in a fast moving style and illustrated by many photographs of the sun and stars.

MOULTON, FOREST RAY: "Consider the Heavens," Doubleday, Doran & Company, Inc., New York, 1935, Chaps. V, VIII, X–XV, inclusive.

The chapters referred to present a thorough but popularized discussion of information about the sun and stars and the use of astronomical instruments.

MASON, FRANCES: "The Great Design," The Macmillan Company, New York, 1934.

This book is a compilation of essays by noted English scientists in which they treat the general orderliness in the universe. Sufficient information is given to establish an understanding of the topics discussed, and then an explanation is made of the meaning and significance of these phases of nature. Many thought-provoking concepts are built up in a language that is understandable and interesting to the nontechnical reader. Some of the essays deal with the stellar universe and the earth as the home of man.

STETSON, HARLAN T.: "Sunspots and Their Effects," Whittlesey House, McGraw-Hill Book Company, Inc., New York, 1937.

The author has presented here in a popularized style the course and nature of sunspots and their various effects upon the earth and human affairs.

DUNCAN, JOHN CHARLES: "Astronomy," 3d ed. rev., Harper & Brothers, New York, 1935.

This excellent and widely used textbook for beginning classes in astronomy in colleges and universities is a good reference for those desiring specific information relating to the various phases of astronomy.

HUBBLE, EDWIN: "The Realm of the Nebulae," Yale University Press, New Haven, 1936.

Scientific research into the outlying regions of the stellar universe has revealed considerable information regarding the nebula which is accurately presented and illustrated with many excellent photographs in this book.

GAMOW, GEORGE: "The Birth and Death of the Sun," Viking Press, Inc., New York, 1940.

A theory of the source of the sun's energy based upon the conversion of hydrogen to helium is clearly presented by the author, a distinguished research physicist.

BARTON, SAMUEL G., and WILLIAM H. BARTON: "A Guide to the Constellations," Whittlesey House, McGraw-Hill Book Company, Inc., New York, 1935.

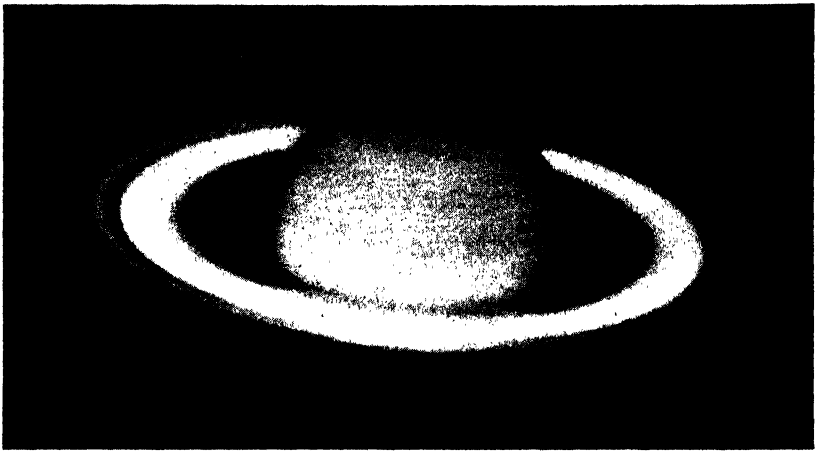
For those wishing a detailed guide to the study of the stars here is an excellent book which contains many maps and photographs.

The Sky, published by the American Museum of Natural History, New York.

This monthly magazine, written for the layman and extensively illustrated, contains articles on a variety of topics in astronomy.

The Astrophysical Journal, published by University of Chicago Press, Chicago.

An international review of spectroscopy and astronomical physics, containing technical articles on research.



Mount Wilson Observatory.

3: THE CELESTIAL FAMILY

A Consideration of the Solar System

THOSE heavenly bodies of keenest interest to man since the beginning of his recorded history are the members of the solar system. The sun has been revered, deified, glorified in poetry, and intently studied by scientists the world over. The moon has been used for centuries to reckon the months of the year, to furnish each new generation with songs, and as the object of many superstitions. The planets have been honored with the names of gods, and many legends were built up regarding them.

Nevertheless these bodies are not of most significance in the stellar universe as a whole. Rather they are relatively so close to the earth that their presence and movements have been more observable and commanding. The solar family consists of nine planets with a number of moons, many thousands of comets, and a vast number of planetoids, in addition to the parent sun which was discussed briefly in the last chapter. All these celestial bodies are so close to the earth as compared to even the nearest star that they are like next-door neighbors.

Their study has revealed much that is useful in interpreting the rest of the stellar bodies. A meticulous examination of our sun has taught us many things that it would have been impossible to learn from the more distant stars. The study of planetary motion has provided us with much of our knowledge of celestial mechanics. All this has been by way of revealing the nature of the universe. In a more practical sense our investigation of the solar system has given us a more accurate method of keeping time, more precise navigation, and a better understanding of weather.

Birth of a Family of Planets

The birth of quintuplets in a human family is an event that attracts the attention of all people, and the details of their rearing are of interest to everyone. The nine planets were, without any reasonable doubt, formed at the same time and under conditions similar for them all. The simultaneous birth of a celestial family of nine (or even fewer) planets is probably a stellar event of much rarer and more remarkable occurrence than the advent of human quintuplets. The question of the earth's formation and likewise of the other planets' has been of great interest to man since ancient times. Many theories, both scientific and religious, have been advanced. No one was here when the earth was formed; no written record was made. Even the earth's rocks give only fragmentary records.

However, with the present methods of mathematical measurements and deductions it is possible to formulate a theory of the formation of the solar system that conforms to almost all the facts of nature so far observed. This is known as the planetesimal hypothesis. It was first advanced by two Americans, Thomas C. Chamberlain and Forest Ray Moulton, respectively geologist and astronomer, at the University of Chicago. A modification of the theory has been proposed by two famous British experts, Sir James Jeans and Dr. Harold Jeffreys, in what is called the tidal hypothesis. This modification involves significant changes in the details of the original theory, but the fundamental idea has remained somewhat the same.

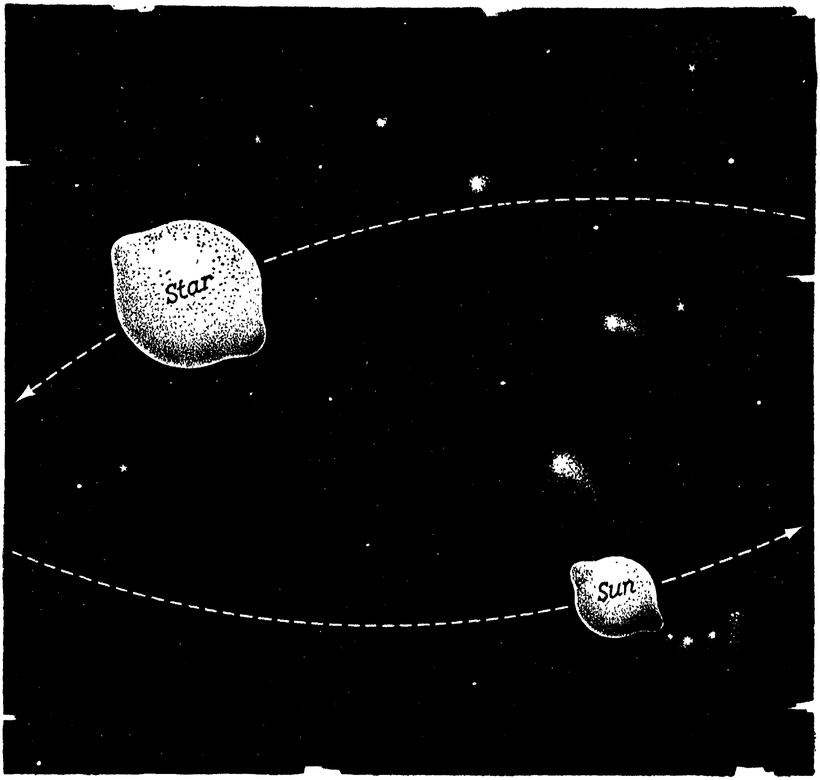
The total hypothesis consists of two essential parts, one having to do with the birth of the solar system and the other con-

cerned with its growth, or development. The formation of the solar system is attributed to the approach of a star relatively near our own sun, probably within a few million miles. The gravitational force of the visiting star caused small parts of the sun at different points to be ejected to such great distances that they did not fall back into the sun but remained outside to form the planets. This is believed to have occurred at least two billion years ago. In those remote times the sun was speeding on its endless journey through space, and the same was true of the other stars. The probability of another star's colliding with our sun is very small, yet it does exist; however, the probability of a star's approaching close to the sun is much greater.

The conditions of the sun then were probably not much different from those existing today. Photographs show that large fountains of matter shoot out from the sun's surface many thousands of miles. The sun has great eruptive possibilities, and it is probable that this has been the case for billions of years. The masses of erupted material ordinarily fall back again because of the sun's gravitational pull; therefore, such material forms no permanent bodies outside the sun. Another force was necessary to pull the erupted masses far enough away so that they would not fall back and to set them in revolution around the sun. This force could have been the gravitational pull of a passing star.

Let it be supposed that the sun and another star do approach and pass each other. The tremendous gravitational attraction of the passing star would be added to the regular eruptive forces of the sun. The result would be great tides on the sun's surface —one on the side directly toward the star and one on the opposite side of the sun. These tides would be the same type as occur today on the earth's seas because of the attraction of the moon. This attractive force plus the eruptive forces of the sun would cause masses of the sun equal to the planets to be drawn out many millions of miles, so far that they would remain permanently outside the sun.

As the disturbing star approached the sun, two gaseous jets, or bolts, would be pulled out from the eruptive material, one on each side of the sun. The one moving toward the star would be larger and would travel much farther than the one on the oppo-



Representing the formation of the solar system.

site side of the sun. This same phenomenon occurs in earth tides, those facing the moon being somewhat higher than those occurring on the opposite side of the earth at the same time. The first larger bolt pulled from the high tide on the sun facing the star was to become the planet Neptune; the smaller one on the opposite side, Mars. A second pair of bolts gave rise ultimately to Uranus and the Earth. A third pair of bolts produced Saturn and Venus, and a fourth pair brought forth Jupiter and Mercury as the visiting star passed out into space.

An illustration of the forces acting on these bolts is shown in the drawing. To explain the formation of the small outermost planet Pluto and the numerous planetoids just beyond the fourth of the minor planets, it has been suggested that they represent small jets of gases ejected from the sun in the early stages as the star first approached near enough to exert an effective pull.

The way in which the earth and other planets have developed after their origin is not quite so generally agreed upon. It is certain that the ejected material cooled down rapidly from the gaseous state; also, that it contained masses great enough to form the nuclei of the planets and that these collected numerous smaller particles and gases. According to the idea of Chamberlain and Moulton most of the matter was in the form of infinitesimal planets revolving around the sun. They cooled first into liquid and finally into solid particles, the central part of each bolt being large enough to collect to itself the smaller particles after they had solidified. The present planets have grown from these small nuclei by gradual addition of the smaller bodies.

The Jeans-Jeffreys "tidal" modification of the theory holds that there was a "dynamic encounter" between a passing star and our sun at a time when the sun was a gaseous mass with a diameter much larger than at present. In other words there was a near collision of the sun and the passing star. Jeffreys has later, however, come to the conclusion that there was an actual collision between the star and the outer regions of the sun and that thereby a great tidal cone was pulled away from the sun. This cone, thought to have been somewhat cigar shaped, consisted of lighter gases from the sun's outer layers at the outer end and denser gases from the sun's deeper layers at the end toward the sun. The outer portion of the cone broke up into the large but less dense outer planets, and the inner part formed the smaller but denser terrestrial planets. According to this theory the planets contained nearly as much mass at the time of their formation as they do now.

The revolution of the planets is nicely accounted for in the Chamberlain-Moulton hypothesis by the cross pull of gravitational forces as the star approached, passed, and receded from the sun. The material ejected from the sun as the star approached would travel in a direction toward the star at that instant. The star is moving, however, and would continue to attract the ejected matter toward itself, thus changing the direction of motion of the ejected material. The result would be that it would follow a curved path rather than a straight one, as shown in the accompanying drawing. The sun is also exerting a force of attraction on the ejected matter and this would further tend to give it a curved path. As the visiting star continued its

journey it would eventually recede to such a distance that its gravitational attraction would cease to affect the ejected material, thus leaving it moving in a curved path around the sun. The same forces would act upon other ejected bolts and give them similar directions of motion as they cooled to form the planets. Once these motions were started, they would continue, with the sun's gravitational pull preventing the planets' escape and causing them to revolve around the sun in elliptical paths. The revolution of the planets is somewhat more difficult to explain with the tidal hypothesis.

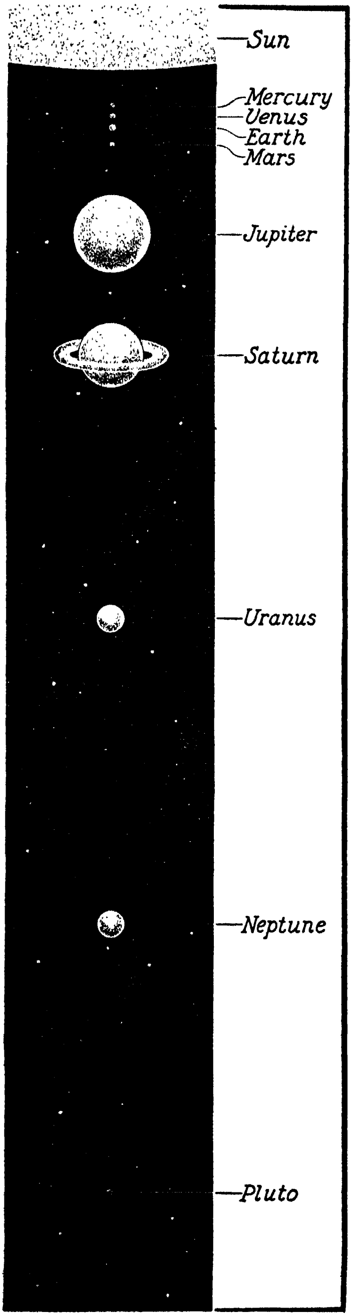
Their rotation on their axes may have been given to them by the impact of particles pulled into them as they revolved around the sun. This collection of small bodies is still continuing with the earth, as is indicated by the thousands of meteors that fall into the earth daily. Most of them burn up when they strike the earth's atmosphere and produce the "shooting stars"; a few of them, however, reach the earth's surface as meteorites.

This is the key to the formation of the solar system as explained by the planetesimal hypothesis, including its modification by the tidal hypothesis. The original hypothesis accounts satisfactorily for the formation of the planets as well as for their revolution around the sun and their rotation on their axes. As such it is generally accepted by scientists.

Resemblances and Differences

The planets have been called the gypsies of the heavens, and from time immemorial they have been known as the moving stars. Of course, every schoolboy now knows that they are not stars. They are not self-luminous, as all the stars are; on the contrary, they are cold and shine merely by reflected sunlight. Compared to stars they are all close to the sun. The nearest is thirty million miles out; the most remote, approximately four billion miles away. In size, they range from a diameter of 3,100 to one of 88,000 miles. Their total mass when compared to that of the sun is but a fraction, it being less than one-seven-hundredth of the sun's mass.

With the exception of Pluto, the planets fall into two groups of four each, according to many of their characteristics and resemblances. The inner, or terrestrial, planets are Mercury, Venus,



Earth, and Mars; the outer ones are Jupiter, Saturn, Uranus, and Neptune. These two groups are different from each other in many ways. The inner planets are smaller than the outer ones. They have a greater density; *i.e.*, volume for volume they are much heavier. However, the total mass of any one of the inner planets is less than the mass of even the smallest of the outer ones, because each of the outer planets is so much greater in volume than either of the inner ones.

The ninth planet, Pluto, is the farthest away from the sun. From what is known at present, it does not possess the characteristics of the other outer planets. It is very small, probably no larger than Mercury, and its mass does not approach the great magnitude of those of the other outer ones. Its differences from the other outer planets might be compared with the differences of the planetoids from the inner planets. (The planetoids are those numerous small bodies which revolve around the sun in an orbit just beyond the outermost of the terrestrial planets.) The relative distances of the nine planets from the sun are shown in the accompanying drawing.

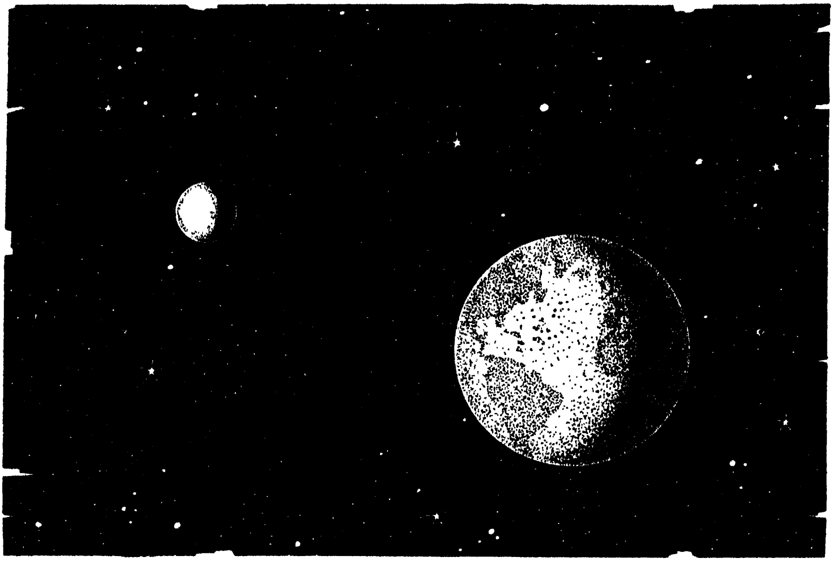
The two groups of planets and their relative distances from the sun. The scale of size of the planets is not the same as the scale of their distances from the sun.

The Inner Circle

The baby planet of our solar family is Mercury; it is nearest the sun. It revolves around the sun in eighty-eight earth days, speeding along its orbit at the rate of twenty-nine miles per second. This swift movement and the fact that Mercury often appears as an "evening star" were the causes of its being named after the Greek god Mercury, the winged messenger of the gods. The planet is believed to have no atmosphere, being so small that its gravitational pull is not great enough to hold any gases on its surface. So far as we know, its period of rotation on its axis is the same as its time of revolution around the sun. If this is so, it keeps one side always turned toward the sun; and this side of the planet therefore reaches a high temperature—about 350°C.

Venus is the sister planet of Earth; in fact it might be called the twin sister, so closely does it resemble Earth in many respects. In diameter it is only 200 miles less, and its distance from the sun is only 25 million miles less. Its year is 225 of our days. It has an atmosphere but is completely and perpetually covered with thick white clouds. Accordingly, its surface has never been seen. Spectroscopic analysis of the planet's clouds fail to show the presence of any oxygen or water vapor. However, the atmosphere does contain much carbon dioxide, indicating that the oxidation of carbon has somehow occurred. The presence of carbon dioxide has made for some speculation that life exists, or has existed at some past time, on the planet, since one result of organic processes is the formation of carbon dioxide.

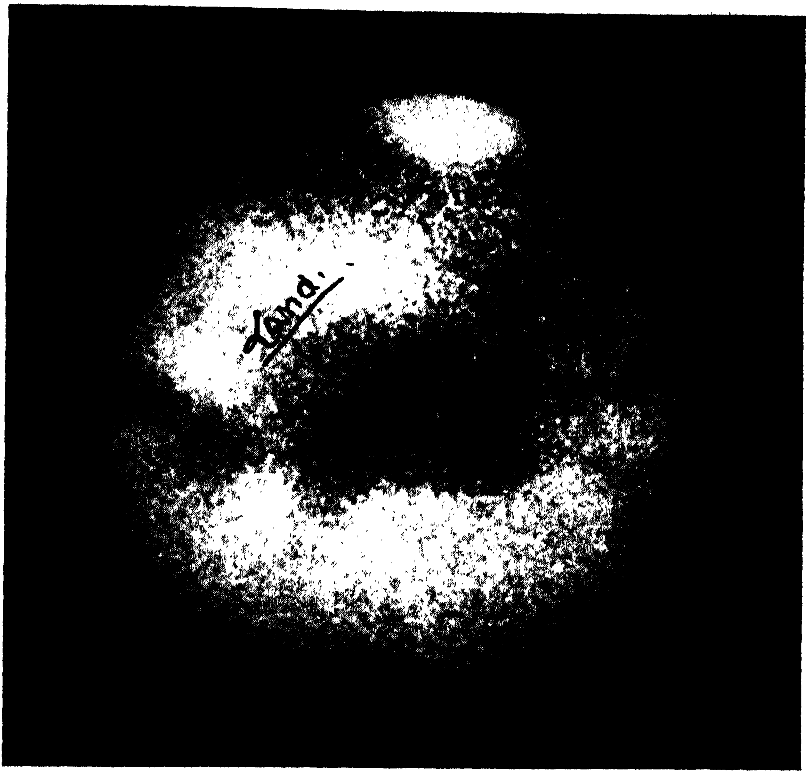
The next planet out from the sun is Earth. If one can imagine himself viewing the earth from another planet, say Venus, it would appear as one of the most interesting objects of the heavens. When the two planets are closest together, Earth would appear as the brightest object in the Venusian sky, except, of course, for the sun. The moon would look like a twin planet to the earth. Since clouds reflect light about three times as strongly as land reflects it, about half the earth would be covered with white patches. The oceans would appear brilliant; forests would be a dull blue; cultivated or grassy lands would have a somewhat lighter color.



The Earth and moon as they would appear from Venus.

Next beyond Earth is the reddish planet Mars, named because of its color after the Greek god of war. Its reflecting power is about the same as that of relatively dark-colored rocks, which indicates that much of its landscape is probably semidesert. Some dark markings appear permanently on its surface; therefore it is possible to measure its rotation with accuracy. The Martian day is only forty-one minutes longer than the earth day; this is remarkable when one considers that its diameter is only a little more than half that of the earth. But more remarkable is the fact that is equator is inclined $23\frac{1}{2}°$ to its orbit, exactly as in the case of the earth. So the climatic zones on Mars are the same as those on the earth; its seasons are the same except that they are longer, as its year is nearly twice the length of ours.

Mars has an atmosphere, and analysis of the light reflected from it shows water vapor and oxygen. However, the air is much thinner than that of the earth, and the oxygen content is only about 15 per cent of what it is in the earth's air. Water vapor is even scarcer, being about 5 per cent of that normally in the earth's atmosphere. Around the poles are white caps, believed to be snow. During their respective winters they increase to about



Mars, showing polar cap and dark areas. (Science Service photograph.)

a third of the distance to the equator, and in the spring they recede. During the Martian summer season the south polar cap sometimes disappears entirely.

Mars is by far the most talked about of the planets. This is because of the peculiar markings, called canals, on its surface and because of speculation as to whether or not life exists there. The markings are thought by some (including the late Percival Lowell, eminent astronomer and founder of Lowell Observatory) to have been planned and constructed by intelligent beings. That such superior creatures exist on Mars is a matter of grave doubt; however, physical conditions there are such that certain forms of plant life and possibly the simpler types of animals may indeed exist.

Once again removed from the sun, are the asteroids, or minor planets. They are tiny objects which form a sort of hazy path

around the sun. They vary in size from 500 miles in diameter down to mere rocks. About 1,500 of them have been discovered and named or numbered. It is quite probable that their number is exceedingly large, most of them being too small ever to be seen from the earth. A theory has been suggested that the asteroids were once a planet and that the latter was smashed to bits by coming too close to Jupiter in one of the planet's revolutions around the sun. Others believe that the asteroids are simply tiny planets that for some unknown reason never united to form a larger planet, as is believed to have been the case with the other planets.

Outer Planets

The giant planet of the solar system is Jupiter, fifth in order of distance from the sun. Its orbit is next beyond the paths of the asteroids. Its mass is about three hundred times that of the earth—greater, in fact, than that of all the rest of the planets together—its volume, about thirteen hundred times as great. Jupiter's surface is continuously covered with clouds which display peculiar belts around it paralleling the equator. The belts are of different colors and change in width and intensity from day to day, often with marked alterations in a few hours. In 1878 a large red spot about 30,000 miles long and 7,000 miles wide appeared in one of the southern dark belts and remained conspicuous for many years. It has appeared and disappeared several times since. No satisfactory explanation has been made to account for it.

When Galileo constructed his first telescope, he discovered four large satellites, or moons, revolving around Jupiter. These are all now clearly visible with small telescopes at times of best observation. Three of them are larger than the earth's moon. Since Galileo's time, seven other satellites have been discovered revolving around Jupiter. Eight of these travel around in the same direction as that in which the planet rotates, but three revolve in the opposite direction. Temperature measurements indicate that Jupiter is never warmer than 100 degrees below zero centigrade. Its atmosphere, then, contains no water vapor or oxygen but only low-boiling-point gases. The clouds probably

The giant planet Jupiter is continuously covered with clouds and surrounded by peculiar belts paralleling the equator. (Mount Wilson Observatory photograph.)

consist of liquid drops or solid crystals of ammonia floating in a cold atmosphere of methane gas.

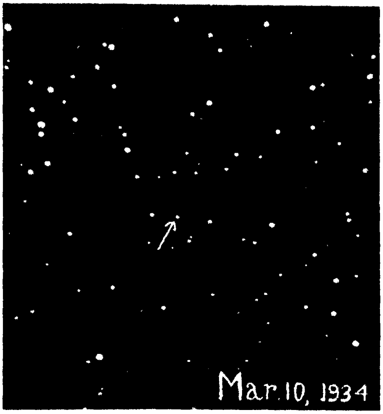
Beyond Jupiter lies Saturn, a planet only slightly smaller than Jupiter. Its large size and the remarkable rings that surround it make it one of the most beautiful objects in the heavens when viewed through a telescope. The rings are three concentric circles of great width but relatively extreme thinness. As they are always viewed by reflected sunlight, they have the appearance of solidarity; yet they cannot possibly be solid. Solid rings would be instantly smashed to bits by the enormous gravitational forces to which they would be subjected. Saturn's rings are composed of small objects, probably meteroic in character, which encircle the planet, as shown in the photograph at the

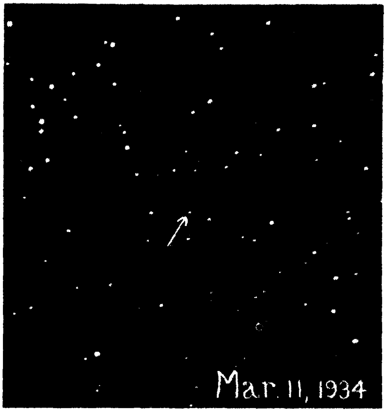
beginning of the chapter. Their origin or significance presents one of the enigmas of the heavens. Since the planet is so far from the sun, it is natural to infer that it is very cold. Such is the case; its mean temperature is approximately 150 degrees below zero centigrade.

Uranus and Neptune are the two other large outlying planets, their relative distances from the sun being in the order named. They are so far from the earth that they appear as very faint stars. The temperature of Uranus is about 180 degrees below zero centigrade, and Neptune is probably 200 degrees below. Both planets are cold, dead objects, with little evidence of activity of any sort taking place on them. It requires 84 earth years for Uranus to make one revolution around the sun and 165 years for Neptune to complete such a journey.

Neptune was discovered before it was ever seen. First its position in the sky was determined after half a century of careful observations and calculations. During this time Uranus was closely watched and was found to have strayed slightly from where it should go according to all the known laws of celestial mechanics. There was only one explanation for this: An unknown planet out beyond Uranus was exercising a gravitational pull. Then, more precise observations and many extensive calculations were made to determine exactly where the unknown planet should be. At last the formulas were solved independently by two young mathematical astronomers, one a student at Cambridge University and the other in Paris. The French astronomer, Leverrier, sent his calculations to the Berlin Observatory. The telescope there was pointed to that portion of the sky, and in less than half an hour the planet was found within a degree of where it had been predicted.

The finding of the planet Pluto in 1930 marked the end of a program of investigation which for unity of purpose and persistence of effort has been matched or exceeded by few other achievements. Following the discovery of Neptune some eighty-five years earlier, many astronomers considered the possibility of finding and locating planets beyond Neptune by their gravitational attraction. One of these astronomers was Percival Lowell. He built the great observatory of Flagstaff, Ariz., which bears his name, in order, among other things, to search the skies





Two photographs of planet Pluto showing its change of position among the stars in one day. (Yerkes Observatory photograph.)

for such planets and to further the study of the planet Mars. The latter was successfully carried out during his lifetime, and a great amount of effort was put on the former. Lowell made a careful study of the orbit of Uranus in order to find some discrepancy that could be used to calculate the position of a more distant planet. In this, however, he was unsuccessful.

After Lowell's death a different program of search for the planet was carried on at the observatory. It consisted essentially of exceedingly careful and persistent photographing of the zodiac, or band of sky that includes the paths of all the planets, and examining these photographs for unknown members of the solar system. About twenty-four years after Lowell's death a young astronomer in the observatory discovered the tiny object's image on one of his photographs. Seven weeks followed in which careful observations and calculations by the Lowell Observatory astronomers confirmed the discovery. The great astronomer's ambition had been achieved although many years after his death.

The period of revolution of Pluto around the sun has been calculated as about 248 years. The distance of the planet from the sun is approximately four billion miles. It is believed to be a little smaller than the earth. Such a small planet at this great distance from the sun would receive little sunlight; accordingly, it is one of the least visible known objects in the sky. The name Pluto means "god of darkness."

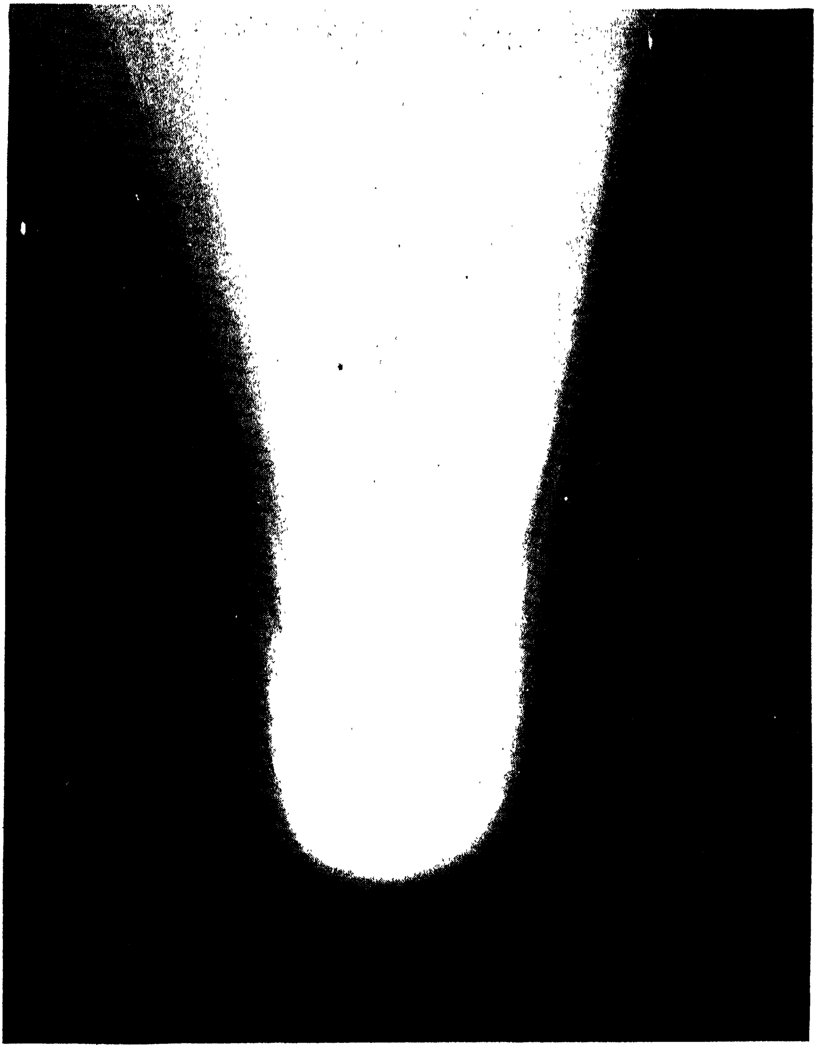
The Comets

The apparently unruly members of the solar family are the comets. They are exceedingly numerous as compared to the planets, and about a thousand have been observed and recorded. Probably there are many thousands more. Often thought of as the fiery bodies of the sky, they have been regarded with fear and superstition. Throughout human history the bright comets have been objects of interest because of their sudden and extraordinary appearance. Their large glowing heads in many cases exceed the sun in diameter, and their streaming tails sometimes extend backward for a hundred million miles, or more than the distance of the earth from the sun.

The source of illumination of the comets is the radiant energy that they receive from the sun. Part of this energy is reflected, much as the moon reflects sunlight, and a part of it is absorbed by the comet and reradiated as light. In most cases the head consists of a nucleus, surrounded by a large envelope of rarefied bright material. The nucleus is not solid but very likely consists of scattered materials of the same nature as meteorites; when it is clearly visible, however, it gives the comet's head the appearance of a starlike eye.

The tail of a comet is a luminous, very rarefied gas, probably no more dense than the best vacuum that could be produced on the earth. One unusual thing about it is that it does not follow the comet continuously. It appears only as the comet approaches the sun; also, it is always on the side of the comet nearly opposite the sun. Various studies have shown that the tail streams away from the head. Evidently it is a part of the large envelope surrounding the nucleus which is pushed back in a direction opposite the sun as the comet moves around that body. This repulsion is due in part to the radiant pressure of the sun's rays when the comet comes near enough for such pressure to become effective and partly, no doubt, to electrical forces within the nucleus caused by ionization of its particles by the action of the sun.

It is now known that no noticeable effect is produced on the earth when it passes through the tail of a comet. On two or three occasions the earth has passed through a comet's tail without any



Head of Halley's Comet, May 8, 1910. (Mount Wilson Observatory photograph.)

ill effects. The last time was in 1910, the year of Halley's comet. No effect was observable except that on one occasion there seemed to be a slight iridescence of the daytime sky.

Many of the comets are known to return at periodic intervals to the vicinity of the sun. Their paths are closed curves, most of them being in the form of elongated ellipses. Halley's comet was the first to be proved periodic when it was estab-

lished by its first definitely known return in 1758 that it had a seventy-six-year period of revolution. Since that time the comet has appeared regularly during the years predicted, and by examining descriptions from past ages it has been possible to identify every return of this famous heavenly body since before the time of Christ. The elliptical paths of the periodic comets vary greatly in their size. The smallest are those which have one of the foci at the sun and the other about as far away as the planet Jupiter. The largest known orbits of the comets extend much farther from the sun than the most distant planet. Likewise, the periods of the time for the different comets to return to the sun differ enormously, ranging from a little more than three to a few thousand years. Some of the comets whose visits to the sun have been observed only once appear to travel in orbits that are either parabolic or hyperbolic, *i.e.*, open curves. If observations and calculations are true, these particular comets are stray bodies from outside our solar system which visit the neighborhood of our sun only once.

Comets are believed by many astronomers to be a second family of the sun. This is particularly true of the periodic ones, and the most recent evidence seems to indicate that a great majority of them are periodic. Just how the comets were formed is less understood even than the formation of the planets. Some hold that the sun picked up this material in passing through a nebulous cloud at one time in its history. Others advance the theory that comets are materials ejected in some unknown manner from the sun. Still others say that they consist of cosmic dust pulled to the sun by the latter's gravitational force. The comets still remain to a considerable degree a mystery, but they no longer need to be regarded with fear or superstition.

The Earth as a Heavenly Body

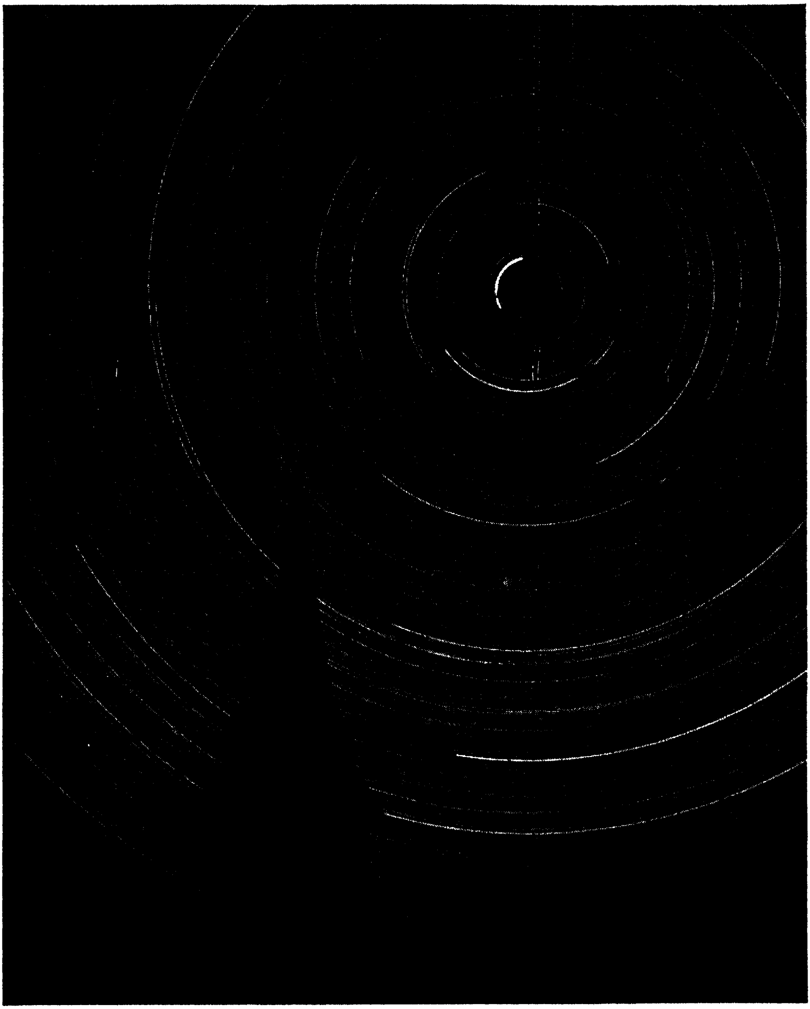
A composite picture of the universe gives us a clearer perspective of the earth as a heavenly body. The known universe is made up of a large number of galaxies, or island universes, separated from each other by distances of the order of a million light years. Each of these island universes is composed of many stars, the numbers of which range from many millions to several billions. One of these island universes is the Galaxy, or our

nebula; it, too, consists of several billion stars, of which our sun is one. The sun is about the size of an average star and similar in character. It possesses a family of nine planets, of which the earth is one of the smallest. As an astronomical body it is seen that the earth is utterly insignificant in the universe as a whole; it is a mere speck of dust in the cosmos. But to people living there, it is the most important of the heavenly bodies, since it provides a home for life.

The motions of the earth in the sky produce conditions that greatly affect life here. Their study, therefore, is one of the most vital and significant of any branch of astronomy.

The earth is a sphere with a diameter of about 8,000 miles. Today this statement is accepted at its face value. But for a long time man believed the world to be flat, and it was hard to convince him that it was round. One of the movements that the earth sphere has is rotation on an axis through its center. This statement, too, is generally accepted today, but man fought the idea bitterly for many centuries. He would have it that the earth remains still and the sun and stars move around it. Nevertheless, we now possess evidence that the passing of the sun overhead by day and of the stars by night is only apparent and that actually the earth is rotating on its axis, thus making them appear to move.

The axis of the earth would, if extended far enough, pierce the sky at two points, one in the Southern and the other in the Northern Hemisphere. The place in the sky to which the north end of the axis points is marked by a bright star, easily identified and known as the Pole Star. The Pole Star would not appear to rotate as the earth turns on its axis beneath the dome of the sky, but all other stars in the Northern Hemisphere would follow circular paths during their twenty-four-hour circuit of the heavens with this star as the center of the concentric circles. When a camera is pointed at the North Star and left exposed throughout a good part of the night, the images produce star trails that show the arcs of these circles, each star producing its own star-trail arc. One interesting point shown in such a photograph is that the Pole Star itself describes the arc of a small circle. It is not, therefore, exactly above the earth's North Pole but a fraction of a degree off. The star-path circles become larger



This artistic picture shows star trails made by pointing the camera directly at the Pole Star and exposing the film continuously for eight hours during the night. (Photograph by J. C. Smith, Colby College.)

as the distance from the Pole Star is increased and would become a straight line for a star directly over the equator.

Another motion of the earth is its revolution around the sun. One of the most noticeable astronomical phenomena is the apparent annual motion of the sun among the stars; *i.e.*, the sun seems to advance continuously eastward among the stars

from day to day, and to make a complete circuit of the heavens in one year. So far as obvious appearances are concerned, it makes no difference whether this fact is explained by the sun's revolving around the earth once each year or by the earth's revolving around the sun within the same period. That the earth does the revolving, however, is shown by three phenomena which may be observed with accurate telescopes. These are the aberration of light, the annual shift of the lines in the spectrum of the stars, and the annual parallax of the stars. These phenomena will not be discussed here, but they are fundamental observations made continually by astronomers.

Telling Time

The aforementioned two movements of the earth are used to measure time. Should you turn to your telephone and dial MEridian 7-1212 (that is, if you live in New York City), you would presently hear a pleasant voice say, "When you hear the signal, the time will be. . . . " Then follows a statement of the time, accurate to within a few seconds. It probably has occurred to only a few that such accuracy necessitates a precise method of reckoning time. Measurement of time is based upon the number of instances of equal duration that occur within a given interval. There is no other way to measure it. The movements that are repeated continuously and with the greatest regularity are those of the earth. These, being subject to man's measurement, are used to measure time.

As everyone knows, the interval of the earth's rotation is a measure of one day's time, and its revolution measures one year. However, a reference point must be taken for these movements, and it must be as nearly fixed as possible. We know that the best such points available are the stars.

To measure time we observe the apparent movement of the stars westward across the sky at night, for this is the best measurement of the rotation of the earth on its axis. An imaginary line in the sky is visualized passing through a point directly overhead and through the Pole Star; this line is referred to as the celestial meridian. A particular star is then selected, and the exact instant noted when it crosses the meridian on its westward journey. The next night this same star will cross the meridian again,

and its exact instant of crossing will again be noted. The interval of time from one transit of the star across the meridian to the next transit is the time that it takes the earth to make one complete rotation. This is called the sidereal day. The sidereal day is divided into twenty-four sidereal hours; each hour is divided into sixty sidereal minutes; and each minute into sixty sidereal seconds. The adjective "sidereal" means pertaining to the stars; *i.e.*, the sidereal day is a unit of time measured by using a star as a reference point.

However, we gauge our activities by the rising and setting of the sun rather than by that of a given star. In practice, the passage of the sun across the meridian is observed instead of the passage of the star. The interval between two successive crossings of the sun is the solar day. It is divided into twenty-four solar hours; each hour in turn is divided into sixty solar minutes; and each of these into sixty solar seconds. Here, the solar day is a unit of time measured by using the sun as a reference point.

This would be a simple procedure so far as measuring time is concerned except that the corresponding units of sidereal time and solar time are not the same length. The sidereal units may be more accurately measured, yet obviously the solar units are the more practical to use. A difference in length between the sidereal day and solar day is produced by the very reference points used to measure the interval of rotation of the earth. As the earth rotates on its axis day after day, it also revolves around the sun, making one complete circuit in one year. However, the earth does not revolve around any of the stars, since they are much too far away. Therefore, the earth will have any given part of its surface facing the sun one time less in the course of one year than the same part will face any star during the year. Accordingly, in the course of one year we have one less day by solar time than there are sidereal days.

The understanding of this condition will probably not be clear to anyone from the mere reading of the foregoing statements. It is rather difficult to visualize from a written explanation that the earth must necessarily make one less rotation with respect to the sun than it makes with respect to a star in one year; however, anyone wishing to verify the fact may do so quite easily with three oranges and a pencil. Insert a pencil

through an orange so that the latter may represent the earth rotating on its axis. Place another orange near by to represent the sun, and another at a greater distance to represent a star. Then move the earth orange one time around the sun orange to represent one revolution of the earth, or one year, at the same time rotating the earth orange on the pencil axis ten times with reference to the star orange; it will be observed that the earth orange has rotated nine times with respect to the sun orange. The conditions of one less rotation with respect to the earth orange than exists with respect to the sun orange will hold true for 365 or any other number of rotations per revolution. When these relative motions are understood, it is easy to see that each of the solar days is slightly longer than a sidereal unit. To be more exact, each solar day is about four minutes longer than the sidereal day.

Another difficulty with the solar day is that it is not uniform in length throughout the year. We are not referring to the different lengths of sunlight or night as one season follows another but to an actual difference in the length of the twenty-four hour solar day. This is due to the fact that the earth travels at slightly different speeds at different places in traversing its orbit around the sun. The sun thus appears to change its speed from month to month as it crosses the sky. In addition, the plane of the earth's equator is tilted at an angle of 23½ degrees to the plane of the earth's orbit around the sun. The net effect is to cause the sun to appear to cross our equator at the beginning of spring and fall, and this northerly and southerly motion tends to retard its easterly journey among the stars. We observe such an easterly journey because the earth is revolving around the sun, and we see the sun against the more distant background of the stars. These two factors—the earth's irregular speed in its orbit and the sun's apparent crossing of the earth's equator—combined make the twenty-four-hour solar day longest on Dec. 22 and shortest about Sept. 17. For example, the sun is moving most slowly across the sky about Dec. 22 when the twenty-four-hour solar day is longest, and it is moving fastest about Sept. 17 at which time the solar day is shortest. Accordingly, a correction must be made, by taking the average of the various speeds and from the average calculating mean solar

time. This is in effect comparable to setting up a fictitious sun in the sky that travels at a uniform speed, since we cannot change the actual speed of the earth, as a jeweler would regulate the speed of a watch. Mean solar time is the universal system now used for reckoning time.

It should be kept in mind that the above-mentioned corrections are small and that highly accurate timing devices and skilled observations are required in order to detect and record them. The value of such corrections lies in being able to measure time with the greatest degree of precision. So far as the daily activities of most people are concerned, it is not important that this fineness of measurement be made. One or two minutes' difference in length of different days would not be significant. Since time is a factor in so many of the happenings of nature, however, it is highly important that it be measured accurately. Speed of moving objects involves the element of time; and if speed is to be measured accurately, the time units used must first be exact. In measuring the distance between Earth and the stellar bodies the speed of the earth as well as the speed of light is involved. Since one of the very few constants of nature is the speed of light, it is a highly important factor, and its exact measurement is of great scientific value. It is 186,284 miles per second. Of course the second must have been accurately determined and found to be of unchanging value before any satisfactory measurement of a speed as great as that of light can be made.

Mean solar time at any point on the earth is called local time; however, local time is not satisfactory, for it is inconvenient to have a different time for every locality. To correct this the railroad companies in the United States in 1883 divided North America into belts, each fifteen degrees of longitude wide. To each belt was assigned the same time throughout, which was the local time of one of the meridians within the belt. There would be, then, no change of time within a given belt. The change in going from one belt to another would be a one-hour jump, since the entire earth would include twenty-four belts. This system was soon adopted by most of the countries of the earth, and it is called standard time. The meridian at Greenwich, England, was taken as the starting, or prime, meridian. Therefore

standard time is now reckoned so many hours east or west of Greenwich.

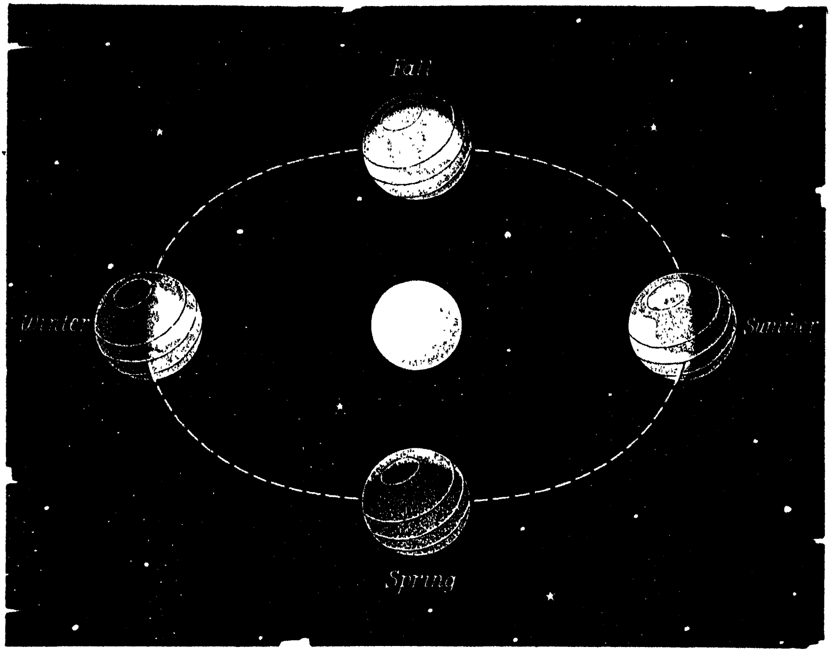
The year is the division of time in which the earth makes one revolution around the sun. This period is not an exact multiple of the day, and there is, of course, nothing that man can do to make it an exact multiple. The earth makes 365.2422 rotations during the time of one revolution. We cannot slow down its rotation or speed up its revolution, so we must correct for the unequal multiple by having a 365-day year and adding one day every four years. The correction, however, makes the seasons come about one day earlier every century, as the fraction is 0.2422 and not 0.25. To make a further correction it has been decreed that there shall be twenty-four leap years every century instead of twenty-five, yet the correction is not quite perfect. The present Gregorian calendar now used in the United States is one arranged to give additional corrections. Providing for an average year of 365.2425 rather than 365.2422 mean solar days, it reduces the error to one day in 30,000 years, and is quite satisfactory for practical and most theoretical purposes.

Seasonal Changes

As the earth hastens on its long journey around the sun each year, one decided effect is produced—a change in the seasons. Many people think that the seasons are caused by the earth's being nearer to the sun in summer and farther away in winter; but this is not true. Actually, the earth is about three million miles nearer to the sun in winter in the Northern Hemisphere than it is in summer; so there is another explanation.

While revolving around the sun the earth is at the same time spinning on its axis, and the gyroscopic effect of the spin is to keep the earth's axis pointing always in the same direction (with only minor variations). The plane in which the earth moves as it revolves around the sun is called the plane of the ecliptic. The earth's axis is not perpendicular to this plane but is inclined at an angle of 23½ degrees to the perpendicular; *i.e.*, the North and South Poles are not pointing straight "up" or "down" from the plane of revolution; but the axis is, so to speak, "leaning."

As a consequence, during the time of half the revolution about the sun the earth's North Pole will be pointing away from



Seasons result from an inclination of the earth's axis to the plane of revolution around the sun.

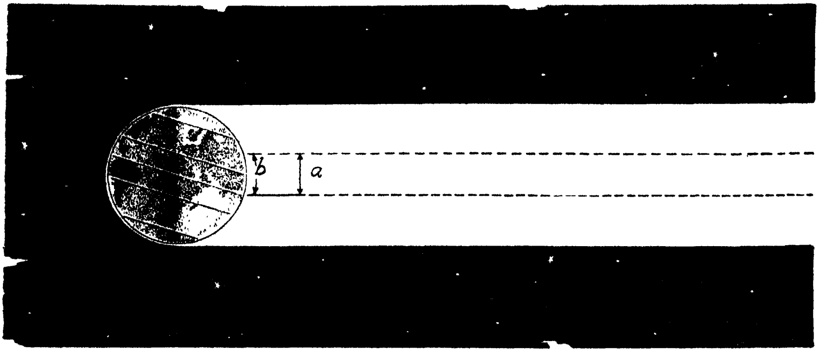
the sun, and during the other half it will be pointing toward the sun. There will be just two days during the year, Mar. 21 and Sept. 23, when the poles are pointing neither away from nor toward the sun. At these two times both the North and the South Poles will be equally illuminated by the sun. Both hemispheres, then, receive the same amount of heat, and they will experience spring and autumn seasons.

While the earth is traveling around the sun from Mar. 21 to Sept. 23, the North Pole is leaning toward the sun and the South Pole is turned away from the sun. The Northern Hemisphere receives more than its normal share of the sun's beneficent rays and is warmer. At the same time the Southern Hemisphere is getting less of the heat and is colder. From Sept. 23 to Mar. 21 the North Pole is leaning away from the sun, and the South Pole is leaning toward it. Then the conditions are just reversed, the Southern Hemisphere having warm weather and the Northern Hemisphere cold.

Let us consider in a little more detail the effect of these conditions on the Northern Hemisphere. From Mar. 21 to June 21 the North Pole is leaning more and more toward the sun each day. The apparent effect of this is to make the sun rise higher and higher overhead toward the North Pole as the days go by. The actual effect is that the sun's rays strike the curved surface of the Northern Hemisphere less obliquely. There is less spreading out of these rays and more concentration per unit area of space. The hemisphere gets warmer and warmer, and we have summer. From June 21 to Sept. 23 the North Pole becomes less and less inclined toward the sun each day, and the concentration of heat received becomes less. On Sept. 23 the North Pole is leaning neither toward nor away from the sun, and both the Northern and Southern Hemispheres receive equal amounts of heat. Then we have fall.

After Sept. 23 the North Pole begins to lean away from the sun, and each day this leaning becomes more and more until Dec. 21. The sun moves across the sky from day to day closer to the southern horizon. As a result the Northern Hemisphere receives less concentration of heat, and on Dec. 21 we have winter. After this date, the North Pole leans progressively less away from the sun each day until Mar. 21, when again the two hemispheres receive an equal amount of heat, and we have spring.

It might be reasoned that the dates given above, June 21, Sept. 23, Dec. 21, and Mar. 21, represent the mid-dates of the seasons summer, fall, winter, and spring, respectively. True, June 21 is the longest day in the year for the Northern Hemisphere, and at that time the sun's rays are coming most nearly perpendicular to the surface, so as to give the greatest concentration of heat; likewise, Dec. 21 is the shortest day in the year, and at that time the sun's rays strike the Northern Hemisphere most obliquely, giving the least concentration of heat. Therefore, it might be thought that June 21 and Dec. 21 would be, on the average, the hottest and coldest days of the year, respectively, and that they would make midsummer and midwinter. They are not the hottest and coldest days of the year ordinarily, because there is a considerable lag in the total heating and cooling of the hemisphere. In summer much of the heat

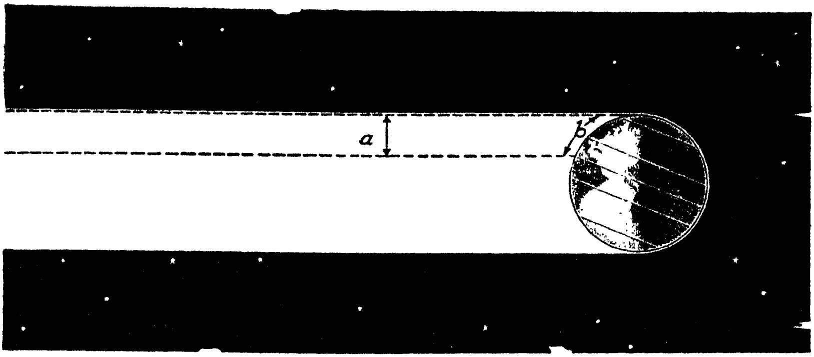


When the North Pole is pointing toward the sun, a given bundle of the sun's rays as represented by (*a*) strike the Northern Hemisphere more nearly perpendicularly and are, therefore, more concentrated on a given area of surface (*b*), producing warm weather.

received is retained for a time. The accumulation of heat tends to make the hemisphere get hotter. The result is that our hottest weather will be a month or so later than June 21. Accordingly, June 21 is ordinarily considered as the beginning of summer. The same conditions apply to the Dec. 21 date. Heat continues to be radiated by the hemisphere after this date at a greater rate than received from the sun's rays, and the weather gets colder for a month or so. December 21 is considered, therefore, to be the beginning of winter.

The leaning of the earth's axis produces an unusual effect near the poles, which is the midnight sun. The land of the midnight sun has been made familiar to us by descriptions of the remarkable effects there as well as by stories of its unique inhabitants. The sun being visible at night is an awe-inspiring sight for those accustomed to the alternation of day and night. It may sink just below the northern horizon for only an hour or two; or it may remain visible continuously and, instead of setting, roll along at "night" from west to east in the northern sky.

Let us see how the midnight sun is produced. When the North Pole is leaning toward the sun, there will be an area around the pole that is continually exposed to the sun's rays as the earth rotates on its axis. In this area the sun continually shines, of course, and does not set. The maximum area so illuminated will be described by a circle around the earth that is 23½ degrees below the North Pole, as may be noticed from the drawings on page 94. At this circle, known as the Arctic Circle, is one day,



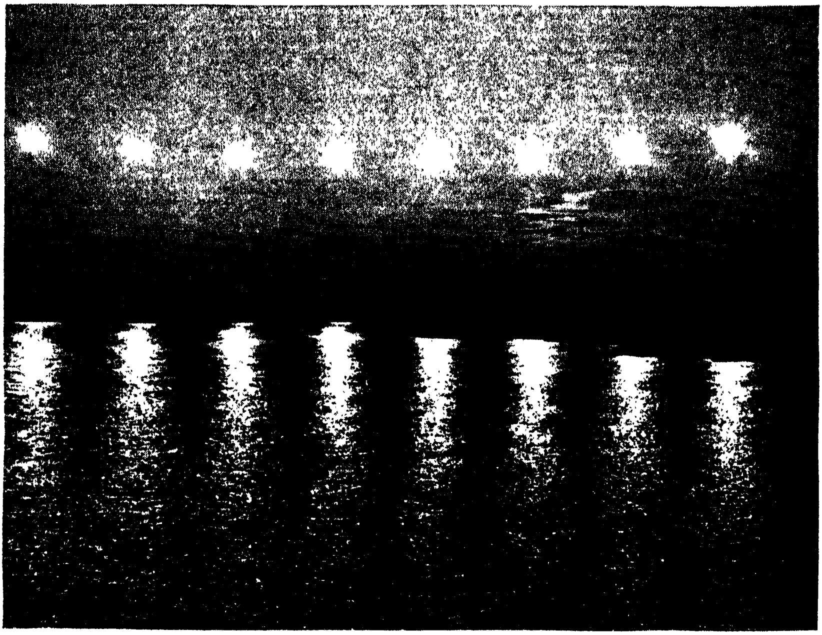
When the North Pole is pointing away from the sun the same bundle of the sun's rays strike the Northern Hemisphere more obliquely and are, therefore, spread over a larger area *b*, producing less concentration on a given area and colder weather.

June 22, when the sun does not set. The number of such days increases as one goes toward the North Pole until the pole is reached, and at the North Pole the sun remains continuously above the horizon for six months from Mar. 21 to Sept. 23.

On Sept. 23 the sun will set at the North Pole, not to appear above the horizon again until Mar. 21. The pole remains in lasting darkness for six months, with the number of such twenty-four-hour days decreasing as the Arctic Circle is approached, where there will be just one day, Dec. 23, when the sun will not rise above the horizon. These same conditions exist at the South Pole, except that the times of continuous sunlight or perpetual darkness are reversed from those at the North Pole.

Earth's Mass

Weighing the earth may seem to the layman like a foolish and impossible task. Archimedes is credited with having said that he could lift the earth if he had a lever long enough and a place to rest it, so great was his admiration for that simple machine. At the same time he might have determined the earth's mass; but unfortunately he did not have such a lever or a place to set it. It is impossible to put the earth on a balance and weigh it. But here, again, the seemingly impossible has been a challenge to man's mind. There are at least two good reasons why man has been interested in weighing the earth. One is his innate curiosity to seek knowledge and truth for the sake of being better informed. The other, of a more practical nature, is his



In the land of the midnight sun, the sun remains above the horizon during the twenty-four hours. The photograph was taken by Donald H. Macmillan near Etah, North Greenland, by exposing the plate every twenty minutes, beginning at 11 P.M., July 25 and continuing until 1:20 A.M., July 26. (American Museum of Natural History photograph.)

wish to establish units for comparison of other heavenly bodies and a more accurate use of the constant of gravity. Knowing the mass and interior structure of the earth has also provided a better understanding of earthquakes.

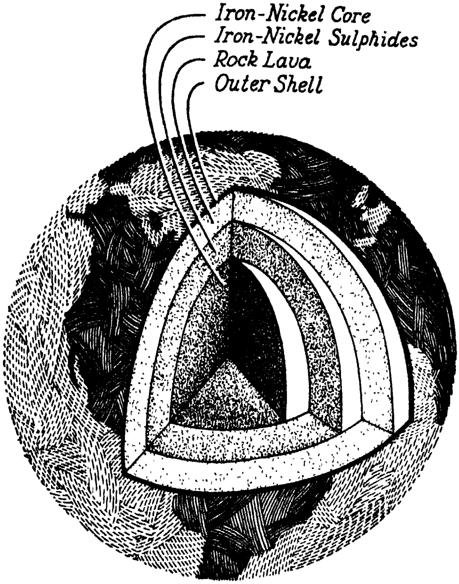
The size and mass of the earth have been measured with great accuracy, and its approximate size is known to everyone. We say that the earth's diameter is about 8,000 miles and its circumference about 25,000 miles. These figures may be obtained with very simple instruments and calculations; they are suitable for general purposes, but as exact knowledge they have little meaning. To make the measurements with great exactness has required enormous efforts and the expenditure of large sums of money. Now astronomers can tell us that the exact diameter of the earth from pole to pole is 7,899.98 miles and that the equatorial diameter is 7,926.67 miles. With these exact figures it has been possible to determine just as accurately the mass of the earth.

The method adopted made use of the law of gravity. We all know that we stay on the surface of the earth as it whirls around on its axis while swinging about the sun. This condition is produced by the earth's gravitational attraction. Perhaps you do not realize that a large, fat man stays on more securely than a small, light man unless you have watched two people of different weight try the pole vault. In other words the force of gravity between two bodies is directly proportional to their masses. For example, the force of gravity between a man and the earth is much greater than the force of gravity between the same man and a soaring airship; a man falling out of an airship will be pulled to the earth rather than held to the airship. The reason is that the mass of the earth is larger than that of the airship and that therefore the pull of gravity of the earth is greater. Use of this knowledge has enabled man to weigh the earth.

To weigh the earth by this process it is first necessary to measure accurately the attraction of gravity between the earth and an object on it. Since distance is a factor in the force of gravity, it is necessary to know precisely how far it is from the center of the earth to the center of the object. Here, of course, it is necessary to recall the exact size of the earth. The third step is to weigh the object. These quantities are then substituted in a well-known equation for gravitational attraction, and the earth's mass calculated.

By far the most difficult task of all was to measure the constant of gravity for the earth, but this has now been accomplished with a high degree of accuracy. Thus, it is possible to make an exact determination of the earth's mass. The results obtained show that the earth weighs some 6,600,000,000,000,000,000,000 tons. This figure, though amazing and interesting, is of more significance because of its size. Since the size of the earth is accurately known, it is possible to determine its density, or weight per unit volume, which turns out to be 5.5 times the density of water. Thus, volume for volume the average weight of the earth is 5.5 times greater than that of water. Now, the heaviest rocks that have been found on any part of the earth's surface accessible to man have a density of only 3.0 times that of water. Most of the materials of the earth's surface are even lighter than this.

Why is the earth as a whole so heavy? The solving of one problem merely produced another one, more interesting and



Interior layers of the earth.

baffling. The unraveling of this riddle not only has shown what is beneath the earth's surface but has had much to do with our present explanations of how the earth had its beginning. The great mass is due to something heavy in the earth's interior. How was it possible to learn this? Obviously there is little possibility of examining directly the inside of the earth. The deepest mines extend down only about three miles, and the greatest tunnels have been drilled through no more than fifteen to twenty miles of pushed-up rock. There must be an indirect way of making the examination.

Oddly enough, the study of earthquake waves passing through the earth, an explanation of the flattening of the earth at the poles and the precession of the poles, an examination of meteors that fall into the earth, and a study of the sun's composition have given us a knowledge of the earth's interior.

As a result of all such studies and the deductions derived therefrom it is now believed that the earth's structure consists of four concentric zones. First, it has a skinlike crust of ordinary

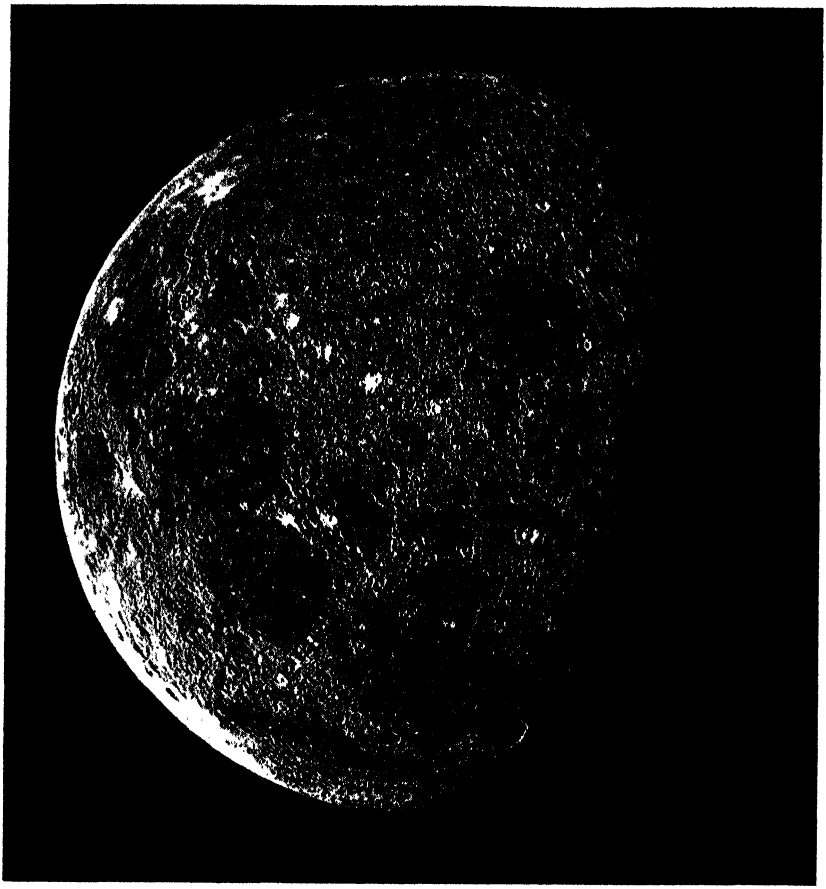
rock, soil, and water about 20 to 100 miles thick. This outer crust is the only part with which man has ever come in direct contact. It forms the continents and ocean beds. Beneath the outer crust lies a shell of lava-like rock 500 to 1,000 miles thick in which "float" the lighter continental rocks and ocean beds. Deeper toward the center is another zone of about the same thickness of iron and nickel sulphides and oxides. Inside this is a core some 3,000 to 4,000 miles in diameter of metallic iron and nickel. These materials constituting the zones would account for the large density and mass of the earth.

Earth's Moon

In addition to being probably the only planet of the solar system that has life of complex forms on it, the earth has one other distinction among the planets. It has the largest moon in proportion to its size in the solar system. If the stars and other planets were all extinguished, our eyes would miss them, and this would be all so far as material effects on the earth are concerned. But if the moon were annihilated, commerce would be greatly affected by the great diminution of the tides. Had it never existed, the development of astronomy and our present understanding of the universe would very likely have been retarded many centuries. The moon owes its significance and economic importance, however, to its nearness, because it is really a very small body as compared to the sun or even to the earth. It is 240,000 miles from the earth, and its diameter is 2,160 miles.

Because of its nearness to the earth, every part of the moon's visible surface, as well as the details of its behavior, has been studied minutely. Viewed through the telescope, it is an awe-inspiring sight. Dusky markings visible to the naked eye are resolved into sharp mountain ranges and level plains. But the most astounding feature is the craters, more than thirty thousand of which have been mapped. Some of them are over 150 miles in diameter and 2 miles deep, and some even have mountain peaks rising from their interiors.

There is neither air nor water on the moon; it is, indeed, a dead world. Any atmosphere or water that might have been there at one time has since escaped, as its gravity is not great enough to hold gas molecules on it permanently. Oceans and



This remarkable relief photograph of the moon shows the moon's face as a chaos of craters and marked by numerous ridges and valleys. (Photograph by Charles H. Cole from a negative made at the Lick Observatory.)

atmosphere are the great equalizers of temperature; lacking these, the moon gets extremely hot during its day and frigid at night. The range of temperature as night follows the day is from the boiling point of water to the temperature of liquid air.

One unusual fact about the motions of the moon is that its period of rotation is the same as its period of revolution around the earth; *i.e.*, it always keeps the same face turned toward the earth. Practically half the moon's surface will remain forever invisible to earth dwellers. The coincidence of its rotation and revolution probably did not happen by mere chance. In bygone

"Old moon in the new moon's arms," indicating clearly why the moon appears to have a different shape at different times. The darkened part is made visible at this time by reflected light from the earth. (Yerkes Observatory photograph.)

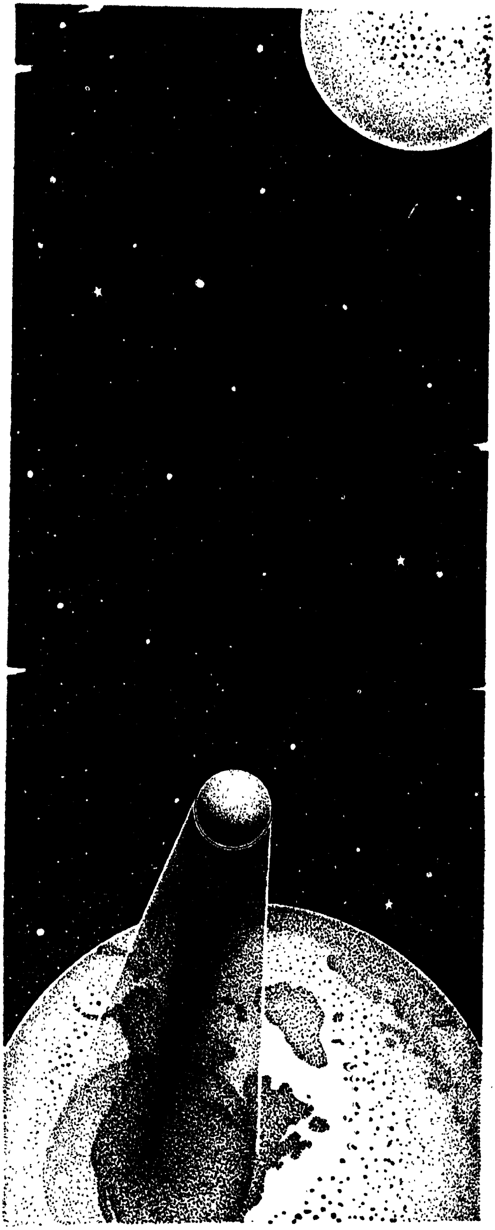
ages when the moon may have been fluid, the earth must have produced tremendous tides upon it. These tides rapidly retarded the moon's rotation on its axis. This retardation would stop only when its day had been lengthened to coincide with its period of revolution around the earth. The tide-producing effect of the moon on the earth is also retarding the earth's rotation and lengthening our day at the rate of one-thousandth of a second per century.

To many people the most perplexing feature of the moon's behavior is the one that is most familiar. This is the change in shape from crescent to full and back to crescent again, as the moon waxes and wanes throughout the month. Obviously, the

moon does not actually change its shape, and the phenomenon is not an eclipse. The answer is simple. The moon is a sphere, and we see it entirely by reflected sunlight. Only half is, therefore, lighted at one time. Looking at the moon from the earth, we may see all of the lighted portion or only a part of it. If we see all the lighted half, the moon will be full, or a complete circle; if we see only a small part, it will appear crescent shaped, since we see only an edge of the half sphere lighted by the sun.

One final glance at the moon before we close this discussion. This is to see its part in giving us eclipses. Total solar eclipses are perhaps the most magnificent spectacles that man is privileged to behold, and for their burst of beauty we owe our gratitude to the moon. An eclipse of the sun comes when the moon gets directly between the earth and the sun. The shadow cast by the moon falls on the earth, and the sun's rays are blotted out. Because of the size of the moon and its distance from the earth, only a small tip of the cone-shaped moon's shadow reaches out to the earth. A total eclipse of the sun, therefore, will be visible only within the small area covered by this shadow. A much larger area will be covered by a partial shadow at the time of each eclipse, as may be evident from the accompanying drawing. Within this area there will be a partial eclipse of the sun, since the sun's disk is only partially covered by the moon, as viewed from any point within this area.

It might be thought that there should be an eclipse of the sun once every lunar month, since during this time the moon passes once between the earth and sun. However, the plane of revolution of the moon around the earth does not quite coincide with the earth's orbit around the sun. It is tipped at an angle of about five degrees. Frequently the moon is north or south of the sun when it passes between the earth and sun. At least twice each year it must come directly in front of the sun; two solar eclipses must occur, then, each year, and there may be as many as five. Most of the eclipses are likely to be partial rather than total, however, at any place where the shadow falls on the earth. On some occasions when the moon gets directly between the earth and sun, it is too far from the earth to cover the disk of the sun entirely. Under this condition the moon at the instant of maximum eclipse will leave a circle of sun showing around the



An eclipse of the sun is caused by the moon getting between the earth and sun and a shadow of the moon falling on the earth.

black silhouette of the moon itself. An eclipse of this type is known as an annular eclipse.

Just what part of the earth the narrow shadow of a total eclipse crosses will depend upon the exact positions of the earth, moon, and sun with respect to each other at the time. A total eclipse was visible in New York City on Jan. 24, 1925. The next one visible in that city will be in January, 1982. When a total eclipse occurs, the period of totality at any one point is at most only seven minutes thirty seconds, usually less. It is a phenomenon, however, that many astronomers travel thousands of miles to view, not merely for its spectacular grandeur but rather to study conditions existing on the sun's surface which are more favorably revealed then than at any other time.

When the moon passes into the earth's shadow, there is an eclipse of the moon. In a manner reminiscent of the case explained above, the earth gets between the moon and sun once each lunar month, but there is not an eclipse of the moon each month. In fact lunar eclipses are not so frequent as solar eclipses. There may be no lunar eclipses at all in a year, and at most there can be only three. At a total lunar eclipse the moon does not get entirely dark but has an unusual reddish glow, caused by the fact that the earth's shadow is not completely dark. The earth's atmosphere bends some of the sunlight passing through it, much as a lens bends light passing through the lens. The red rays are bent most, and penetrate the earth's shadow to light up the moon with the peculiar red halo at the time of its total eclipse. Those who have never seen a total eclipse of the moon will be well repaid for availing themselves of the next opportunity to do so.

REFERENCES FOR MORE EXTENDED READING

WILLIAMS, H. S.: "The Great Astronomers," Simon & Schuster, Inc., New York, 1930.

A well-written book that gives not only many glimpses into the lives of the great astronomers but also a wealth of material relating to the development of the science of astronomy.

JEANS, SIR JAMES: "Through Time and Space," The Macmillan Company, 1934, Chaps. I, III, IV, V.

In the chapters referred to is found an excellent nontechnical account of the earth, moon, and planets as astronomical bodies, illustrated by a few remarkable photographs.

WATERFIELD, R. L.: "A Hundred Years of Astronomy," The Macmillan Company, New York, 1938.

The last century has witnessed in astronomy remarkable discoveries, many of which are dramatically and thoroughly described in this book.

MOULTON, FOREST RAY: "Consider the Heavens," Doubleday, Doran & Company, Inc., New York, 1935, Chaps. III, VI, IX.

In these chapters one of the authors of the now widely accepted planetesimal hypothesis of the origin of the solar system has given an interesting and readable account of the origin of the planets and many facts relating to the solar family.

FATH, E. A.: "Elements of Astronomy," McGraw-Hill Book Company, Inc., New York, 1934.

An excellent textbook for college students of astronomy, and one suitable for the capable general reader who wishes specific information on a given topic.

BARTKY, WALTER: "Highlights of Astronomy," University of Chicago Press, Chicago, 1935, Chaps. I, II, IV, VI.

The chapters referred to are the parts dealing with the solar system of a textbook written for the introductory study of astronomy in the New Plan at the University of Chicago. The content has been well selected and organized for this purpose. The style of writing is interesting and fast moving, and the book contains a number of appropriate illustrations.

MILLIKAN, R. A., *et al.*: "Time and Its Mysteries," New York University Press, New York, 1936, 1940. (Ser. I and II.)

James Arthur bequeathed his unique collection of timepieces to New York University and created a foundation for an annual lecture on "Time and Its Mysteries." The eight lectures by some of America's most noted scientists published in these two volumes present outstanding contributions to the subject of time from several points of view.

RUSSELL, HENRY NORRIS, "The Solar System and Its Origin," The Macmillan Company, New York, 1935.

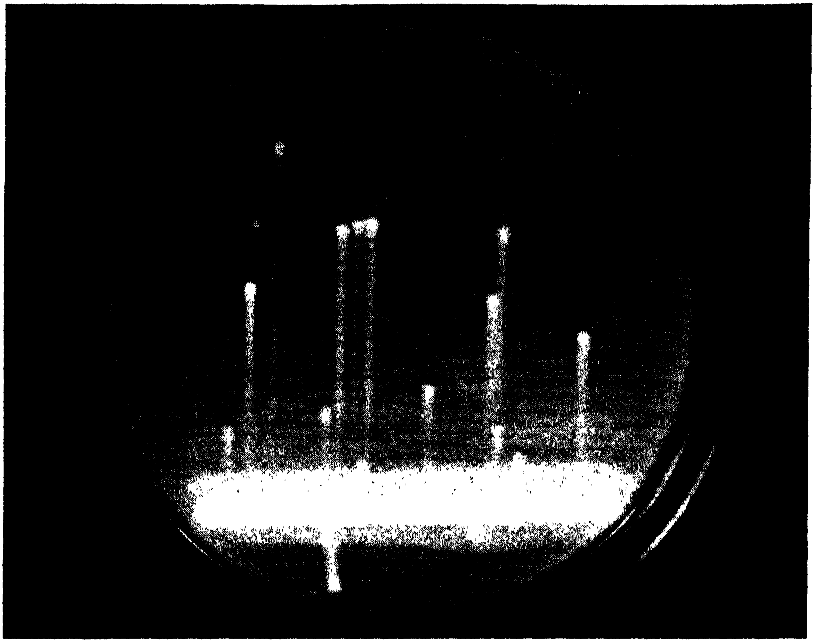
An excellently organized and well-written little book which gives a thorough account of the factors involved in determining the origin of the solar system. It points out specifically some of the defects in modern theories and sets forth certain deductions and conclusions based upon highly refined modern research and mathematical calculations.

Popular Astronomy, published by Goodsell Observatory, Northfield, Minn.

This monthly journal contains many articles that will be of interest to the general reader. They are usually written in nontechnical language and well illustrated.

The Telescope, published by Harvard Observatory, Cambridge, Mass.

The inquiring layman will find in this bimonthly magazine many articles of interest and also a current map of the sky.



Black Star.

4: BUILDING PARTICLES

A Consideration of the Nature and Structure of Atoms

IN MEMORIAL HOSPITAL in New York City is a quantity of radium weighing eight and one-half grams, or about one-third of an ounce, the commercial value of which is approximately $250,000. It is kept permanently sealed in a thick lead chamber. Should it ever be removed, it would be handled behind a sheet of lead, so deadly to living tissue are its rays. From a practical standpoint, radium is the basis of one of the most useful methods known to man for treating cancer. In addition, the behavior of this substance has given us some insight into the fundamental nature of all material things. Compared with other objects of similar size, it is the most valuable and most important material in the world.

For the discovery and study of radium the Nobel Prize, highest honor to be awarded for scientific research as well as for outstanding contributions in other fields, was twice given in one family, first to Pierre and Marie Curie together, and later to Marie Curie alone. Madame Curie is the only one in the field of science to receive it twice. Radium proved to be the touchstone that fostered much of the development of modern science during the early part of the twentieth century. The most important property of radium is that it is continuously disintegrating and giving off particles and radiation; and by its self-destruction it reveals much concerning the fundamental structure of matter.

Until the beginning of the seventeenth century, mankind had little understanding of the structure of the material world. With the exception of a few learned ancient scholars, man was content to believe that stones were stones, fire was fire, and water was simply water. Now we know that all substances are composed of extremely small, invisible particles, called atoms—the veritable building stones of the universe. Billions and billions of them compose the human body. They make up the air that man breathes, the food that he eats, the ground on which he walks; and their interactions furnish the energy that he uses, even the energy of life itself.

A Lesson in the History of Little Things

With this general picture in mind, the question at once arises, What are atoms like? To determine the exact nature of atoms has been, indeed, a baffling problem. For a hundred years some of the best minds on earth have been engaged by it, and today research is being carried out on a wide scale. A brief survey of the historical and present study of atomic structure will reveal how significantly our knowledge in this respect has increased and will also give an insight into the present understanding of the nature of atoms.

The word atom is derived from the Greek meaning "uncut, indivisible." The ancient Greeks were great lovers of knowledge, and few subjects escaped their inquiring minds. Some of these early students of natural science speculated upon the structure of matter and noted that a stone could be divided and further subdivided. This could continue, they reasoned, until the divided

particles were of the fineness of powder, when the limit of divisibility would be reached. The same would hold true for other common substances, such as wood or water or minerals. They called these smallest particles the atoms. However, the Greeks were philosophers and not experimenters, and so the question of the ultimate structure of matter rested with this speculation. With the decline of Greek civilization, any study of the fundamental structure of material things seems to have been cast aside for more than sixteen centuries.

The atomic theory of the structure of matter was first experimentally established by John Dalton, a schoolmaster at New College, Manchester, England, in 1802. He recognized the simple forms of matter as chemical elements and reported that these as well as more complicated forms were made up of atoms. He conceived atoms as being tiny, indivisible particles like small round beads or minute billiard balls with a surface and a definite shape. Dalton's hypothesis concerning the fundamental nature of atoms prevailed in somewhat modified form until the very close of the nineteenth century, and he is often referred to as the "father of the atomic theory."

Atoms were studied for nearly a hundred years by Dalton and those who followed him chiefly by chemical methods. The properties and characteristics thus discovered were those which atoms show in the aggregate, or the large numbers that take part in chemical changes. In chemical reactions, atoms are the smallest units taking part, and it was logical for Dalton and his followers to reason that they were indivisible. Not enough energy is involved in chemical changes to disrupt the atom, and this condition is as true today as it was in Dalton's time.

By the end of the nineteenth century much information regarding the atomic structure of matter and the general nature of the atom had been acquired. Most of the ninety-two chemical elements had been discovered, and it was found that the atoms of each element were different in chemical and physical properties from those of the other elements. For example, it was discovered that all atoms did not have the same weight. These weights, called atomic weights, ranged from 1.0078 for the hydrogen atom to 238 for an atom of the metal uranium, with oxygen taken as the standard and arbitrarily assigned the atomic weight

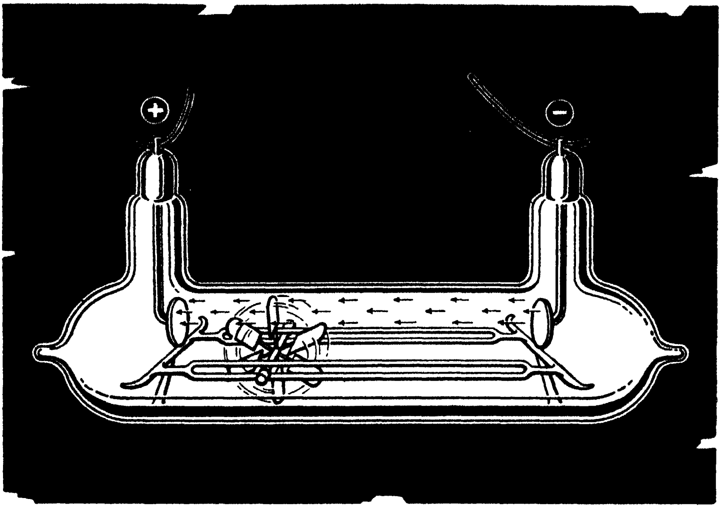
of 16. This particular weight was assigned to oxygen so that many of the elements would have atomic weights close to whole numbers, and none would have an atomic weight of less than one.

A further discovery was that atoms combined in small numbers to form molecules and that in all substances there was a random motion of the atoms and molecules as a whole. The extent of this random movement was found to be a measure of the heat content of the object. Also, the average diameter of the atoms had been determined, and it was of such exceedingly small size that about two hundred million atoms laid side by side would equal just one inch. More specifically stated, the diameter was determined to be of the order of 10^{-8} centimeter, a figure that holds approximately true in the light of more refined present-day measurements. Although this figure represents the average diameter of the atoms, their exact size is different for each of the different chemical elements.

In addition to the above-noted physical properties, certain fundamental chemical characteristics had been discovered. One of these was that atoms could be grouped according to their atomic weights into some eight groups, or chemical "families," the chemical properties of which were similar. For example, the chemical property of the noble-gas family, or Group 0, is that of a very inactive element. In this group are found the so-called inert gases helium, argon, krypton, and xenon. They never enter into chemical reactions and never form compounds. Other chemical families contain very active elements which readily combine with other elements. In one such family, Group 7, are found fluorine, chlorine, bromine, and iodine.

New Knowledge of an Old Topic

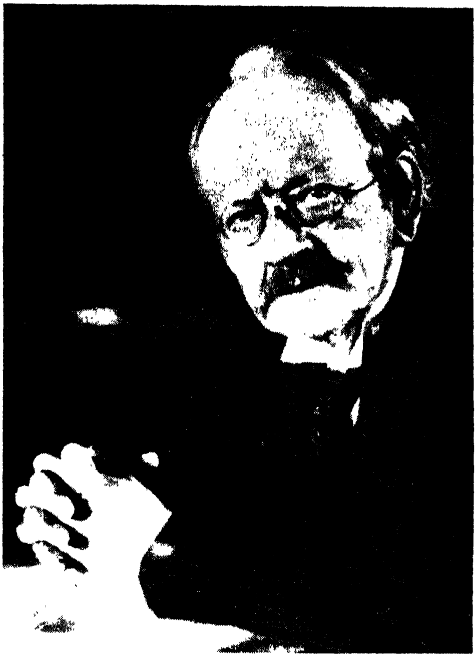
The idea of Dalton that atoms were indivisible units of matter and that the atoms of each of the elements were entirely different entities had become firmly established by the end of the nineteenth century. Then in 1895 a series of discoveries which completely upset this well-established picture followed each other in rapid succession. These discoveries marked the beginning of the whole of modern scientific thought and gave a much more complete and greatly altered picture of atomic structure and behavior.



Electrons moving through a specially constructed vacuum tube cause a paddle wheel to rotate.

The first was made by William Crookes, an amateur physicist of London. Crookes's experiments consisted of passing a high-voltage electric current between two electrodes sealed in opposite ends of a glass tube filled with gases at very low pressures. A beautiful fluorescent light was observed emanating from the negative electrode. Crookes found that if the rays were allowed to fall on an object, they would heat it. He discovered, furthermore, that they had sufficient energy to rotate a small paddle wheel placed in their path, indicating that the light in the tube was produced by material particles, since force was exerted to rotate the paddle, a force similar to that of particles of water hitting against a water wheel. Later he found that these particles could penetrate thin walls of glass or metal. They were deflected by a magnet in such a way as to show that they carried a negative electric charge. For his momentous discoveries Crookes was finally recognized in British scientific circles and knighted by the British Crown.

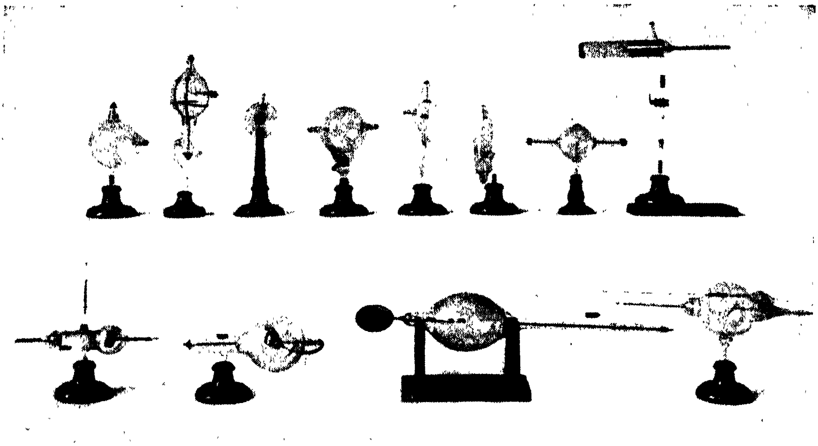
Stimulated by this work, Sir J. J. Thomson in 1897 discovered that the particles from Crookes's tubes were electrons and that they were driven off the atoms of the metal sealed in the tube. He compared the weight of one electron with that of the hydrogen atom, believed at that time to be the lightest



Sir J. J. Thomson. (Science Service photograph.)

particle in the material universe, and found that the former was only about 1/1,800th that of the hydrogen atom. Here was an extremely light particle constituting the negative charge of electricity emitted by atoms heretofore believed indivisible! Atoms were not the indivisible units that they had so long been considered.

Simultaneously with Thomson's work, Wilhelm Konrad Roentgen of Germany was bombarding a metal target (which he had placed inside a Crookes tube) with the electrons, or cathode rays as they have come to be called, that emanate from the electrode of the tube. To his amazement and delight a sensitive screen placed near the bombarded target but outside the glass tube began to glow with fluorescent light. Roentgen was at a loss to explain this phenomenon except that it was produced by a new kind of radiation which had the power of penetrating glass and which was being emitted by the atoms of the metal target when it was bombarded. He replaced the sensitive screen with a photographic plate wrapped in black paper. When the



Various X-ray tubes used by Prof. Konrad Roentgen as preserved in the Deutsches Museum, Munich. (Courtesy of Eastman Kodak Company.)

plate was later developed, it showed that the rays had penetrated the paper and exposed the plate. He further found that these rays would penetrate the soft parts of his hand and that he could use them to make shadow photographs of the bones of his hand and arm.

Roentgen announced his discovery to the German Physical Society in 1895 and on that memorable occasion exhibited to an astonished group of scientists photographs of the bones of his hands and photographs of coins taken through the leather of his purse. Here was evidence of an unknown and penetrating form of radiation coming from the atoms of a metal target when it was bombarded with cathode rays. The apparently well-ordered atomic world was being thrown topsy-turvy. Because of the unknown character of these rays, they were called "X rays" after the unknown quantity in algebra.

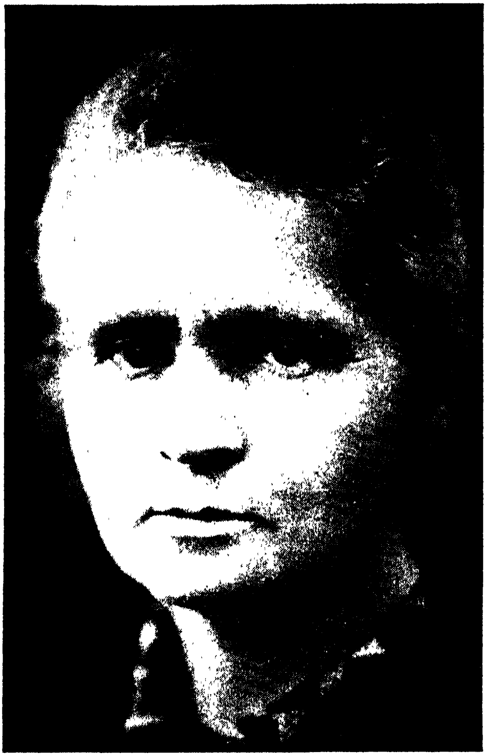
During the following year, 1896, Henri Becquerel, in Paris, discovered that minerals containing the element uranium emitted rays that would affect a photographic plate in the same way that Roentgen's X rays affected it. He made the discovery while searching to find whether or not there were any minerals that would emit the mysterious X rays if exposed to sunlight rather than to the cathode rays of Crookes's tube. He exposed various substance to sunlight and noticed their effect on

photographic plates. During a period when the sun did not shine, Becquerel was impatient to continue and so placed some pitchblende (containing uranium compounds) near a photographic plate in a completely dark room. After a lapse of several days, he found that radiation from the pitchblende had exposed the photographic plate although it had been covered with opaque, black paper. Penetrating radiation had been given off by the uranium without any exposure to the sun or any other outside stimulation. Emanations had been produced by an inherent property of the uranium atoms.

One of the X-ray photographs made by Roentgen, it being first exhibited in 1895. (Courtesy of Eastman Kodak Company.)

Stimulated by Becquerel's findings, two fellow countrymen, Pierre and Marie Curie, set out to determine whether the source of these emanations was the uranium itself or something associated with it. Also, they wished to find if other substances were radioactive, *i.e.*, would emit radiations. They examined several minerals and laboriously analyzed several tons of pitchblende from which they extracted about a teaspoonful of a highly active substance, an element a million times as active as uranium itself. This they named radium. Radium proved to be a fascinating substance in its radioactive properties, and with it the study of the nature of atoms was accelerated manyfold.

Further experiment with radium and uranium showed that three unique emanations were actually coming from radioactive substances: First, some of these emanations would penetrate only thin paper or thin glass, and they were deflected by a mag-



Madame Marie Curie. (Science Service photograph.)

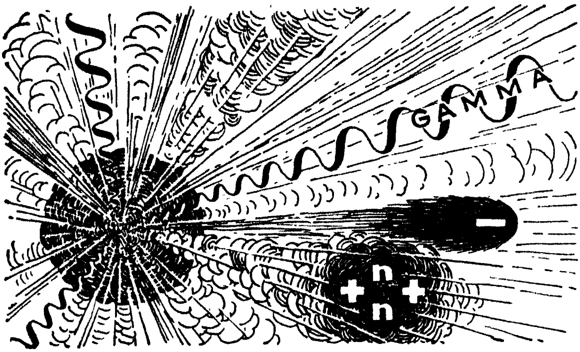
net in such manner as to indicate that they carried a positive electric charge. Second, some would penetrate an iron plate and were shown by a magnet to have a negative electric charge. Third, the most penetrating rays were not affected at all by a magnet and they would expose a photographic plate. These three kinds of rays emitted by the exploding atoms were soon named, respectively, alpha, beta, and gamma rays after the first, second, and third letters of the Greek alphabet.

Sir Ernest Rutherford, a brilliant English physicist, began a study to determine the nature of the alpha rays, the least penetrating of the emanations coming from radioactive substances. By means of a unique but simple little device, he was able to collect in a glass vial the alpha rays from radium. They were then analyzed by means of the spectroscope and found to be helium gas. It was evident that the alpha rays were a

Sir Ernest Rutherford. (Science Service photograph.)

stream of helium atoms moving with large energies. However, since they had a positive charge, they were not normal helium atoms but rather helium atoms stripped of their negative charges, or electrons. They are now known to constitute the nuclei of helium atoms. Furthermore, it was soon discovered that the particles making up these rays each had a mass of four times the weight of the hydrogen atom and a positive electric charge of two. They are therefore relatively heavy atomic particles sent out of the radioactive atoms like bullets from a rifle and ejected with a velocity of about ten thousand miles per second.

The beta rays were investigated chiefly by Sir. J. J. Thomson of England. They were found to be a stream of particles much lighter than the alpha particles, being 1/1,835 the mass of the hydrogen atom, and hurled from the radioactive atoms with terrific velocities which ranged from 60,000 to 180,000 miles per second. Beta rays are more penetrating than alpha rays; also, they are deflected by a magnet in such way as to show that they have a negative electric charge. The beta rays were found to be the same as those produced by Crookes in his vacuum tubes;



Three "rays" emanate from radioactive atoms.

i.e., these strange particles coming from the atoms were nothing more than high-speed units of negative electricity, the electrons.

The third type of rays coming from the radioactive materials were found upon investigation by a number of experimenters to be not a stream of particles but rather a form of radiant energy, similar to light waves and the X rays discovered a few years earlier by Roentgen. They have a higher penetrating power than either the alpha or the beta rays, and they are not deflected by a magnet. They are the gamma rays.

In brief, when radioactive atoms explode, three "rays" come out. Only one type, the gamma rays, is a form of radiant energy. The other two types are particles. The beta particle is a high-speed electron, the unit of negative electricity; the alpha particle is the nucleus of the helium atom.

It was soon discovered that these particles are not simply held captive in the radioactive materials, being allowed to escape with these high velocities, but that they come from the very nuclei of the atoms of the radioactive materials. This means, of course, that helium gas is produced from exploding radium atoms. Here was a realization of the dream of the alchemists of old, *viz.*, the possibility of converting one element into another by transmutation. Furthermore, it was found that when the radium atom explodes and emits helium nuclei and electrons, it is no longer a metallic radium atom but rather is changed to the atom of a gas known as radon. Radon is itself radioactive and soon breaks down, emitting alpha rays and forming another kind of radium, called radium A. From the point of view of

chemistry it is significant that the radioactive gas radon falls into Group 0 of the Periodic Table and is chemically an inert gas with chemical properties much like those of neon or argon. Radium A is also a radioactive substance and disintegrates into another element. This transmutation continues for several other steps until finally lead is produced. Lead, which is not radioactive, is the end product of all radium disintegration.

The apparently simple and indivisible atom of Dalton has been shattered to its foundation by these discoveries. Atoms were thought to be complex aggregates of still smaller and electrically charged particles. As the result of these discoveries an entirely new picture of the composition and structure of matter was produced. The concept or reality of the atom has not been destroyed in any sense; on the contrary, it has been more firmly established. Nevertheless, by 1929 discoveries and mathematical theory had produced a picture of an atom with an entirely different anatomy from that assumed by Dalton and his immediate successors.

Nucleus and Surrounding Particles

On the basis of the discoveries and mathematical calculations made up to 1929, it has been definitely established that atoms consist of a compact nucleus surrounded by one or more negative electric charges. The particles within the nucleus contain the positive electric charges and most of the mass (or weight) of the atoms, and they are tightly squeezed into a relatively small space. Outside the nucleus are scattered one or more electrons which revolve around it at great speeds. These are usually called the extranuclear electrons, and they occupy the greater part of the volume of the atom.

For example, the hydrogen atom, the lightest and simplest of all the atoms, is known to be composed of a nucleus that consists of a single, relatively heavy particle with a positive charge. Therefore, the unit positive charge of electricity was conceived as the nucleus of the hydrogen atom and was known as the proton. Surrounding the nucleus of one proton in the hydrogen atom is one negative electric charge, or one extranuclear electron. Likewise, atoms of other elements were thought to be built up of combinations of two or more protons and electrons.

It seemed definitely established that protons and electrons were the fundamental particles from which all atoms were constructed; hence, all matter was electrical in character. The whole of the material universe was considered to be composed of electricity and nothing but electricity. This concept seemed to explain almost but not quite all the characteristics of the atoms that had been experimentally discovered at that date.

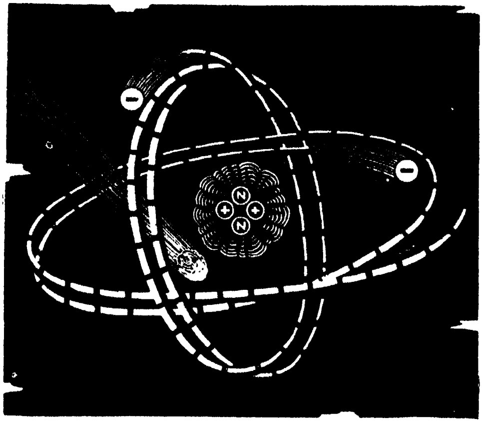
Furthermore, the mass of the proton had been accurately determined and was found to be 1,835 times as great as that of the electron, making its mass equal to 1.0072 atomic weight units. In these units the mass of the entire hydrogen atom is 1.0078, as has been previously mentioned. The proton, therefore, was almost the entire weight of the hydrogen atom, and the remainder was the weight of the outer electron. With respect to size, the electron was found to be so small that about fourteen trillion electrons laid side by side would equal one inch. The diameter of the electron, therefore, is about 2×10^{-13} centimeter. This, it will be recalled from the figure given on page 24, is about fifty thousand times smaller than the average atom, if it is possible to visualize anything so small. The proton, or nucleus, of the hydrogen atom was calculated to be smaller than the electron, approaching the infinitely small limits of approximately 1/2,000 the diameter of the electron, or about 10^{-15} centimeters. Thus, the proton was an exceedingly small, compact unit, whereas the electron was a much larger although lighter unit.

The foregoing sizes of the electron and proton were determined by measuring accurately the size of the electric charge of the electron and basing all calculations upon the electromagnetic theory of the origin of mass. This is neither the place nor the time to discuss this theory, but suffice it to say that within recent years doubts have been cast upon it. However, from other evidence having nothing to do with the electromagnetic theory, it is now certain that the excessive minuteness of both the proton and the electron as predicted by the theory is correct, although we have no way of knowing what the exact sizes are.

Open Structure

It must be quite evident from the preceding discussion that the atom is, therefore, an exceedingly open structure, the very

small, compact nucleus being at the center and the relatively few extranuclear electrons occupying only a small fraction of the total volume. Calculations have shown that 99.999999998 per cent of the apparently solid copper atom is empty space. Approximately the same thing is true with other kinds of atoms. If these dimensions are difficult to visualize, and of course they are, imagine the atom as magnified with a microscope of the mind until it approaches the size of the solar system. The particles inside would then compare somewhat with the planets in size. The whole dimension of the atom would be as empty as is the solar system.



A high-speed atomic particle passing through the space within a helium atom.

It is rather common knowledge that high-speed atomic particles by the billions may be shot through a piece of glass without altering the characteristics of the glass in any respect and without leaving any holes in it. Alpha particles pass through the open latticework of the atoms of glass just as a comet passes through the solar system without striking the sun or a planet. It is not difficult to understand, then, how high-speed atomic bullets such as alpha or beta particles may penetrate sheets of apparently solid metal or wood or human tissue. So the atom was visualized as a sort of "solar system model," with a small, heavy nucleus at the center surrounded by the outer electrons, and was compared to the sun as the center of the solar system with the planets revolving around it.

Thus again man had constructed a complete and fairly satisfactory picture of the nature of atomic structure and behavior, based upon the facts at his command. However, incomplete knowledge is usually more or less inaccurate, and in 1930 a series of discoveries was begun that has required certain altera-

tions in this well-established idea of atomic structure. This new picture tells us more about the nucleus of the atom and gives a different view of the extranuclear electrons constituting the outer part of the atom.

Deeper into the Atomic Nucleus

From 1930 to 1933 there were performed in France and England a series of brilliant experiments which again produced some fundamentally new ideas regarding the nature of the nucleus of the atom. The names that came into world prominence in the field of science as a result of these discoveries were those of Frederic Joliot and his wife Irene Curie, daughter of the late Pierre and of Marie Curie, in France; and that of James Chadwick of Cambridge, England. Without attempting to outline these experiments, it is sufficient here to note that a different kind of particle from any previously known had been discovered coming out of the nucleus. It was relatively heavy, of about the same mass as the proton, but with no electric charge. Here was mass without electric charge, the first such unit that had been discovered since the advent of highly refined modern analytical techniques. It was called the neutron. A particle of mass entirely dissociated from electric charge coming from the nucleus of the atom was such a revolutionary discovery that at first it was greeted with skepticism. The existence of the neutron within the nuclei of atoms was soon verified, and again it became necessary to revise the ideas regarding the structure of the atom.

Recent studies have given some information regarding the neutron. Its properties are in many ways like those of gamma rays. In fact for some time scientists believed that neutrons were merely high-energy gamma radiation. However, it now appears that the neutron is a small, compact particle of about the same size and mass as the proton. Its mass has been measured by various methods; for it must be remembered that neither the neutron nor any other atomic particle can be put on a balance, however delicate and accurate, and thus weighed. Determination of the mass of atomic particles is accomplished by deductions from other accurately observed reactions. One of the most accurate of these reactions has been to produce neutrons and

protons by the disintegration of the nuclei of stable atoms, using gamma rays to supply the energy for disintegration. The latest measurement of the energies involved in these disintegrations shows the mass of the neutron to be 1.0092 as compared to 1.0072 units for the proton on the atomic weight scale.

Beyond knowing that the neutron carries no electric charge and knowing its approximate mass, there is no certainty at present as to just what it is. It may be an entirely distinct and fundamental unit of mass without any vestige whatever of electric charges; and all its characteristics that it has thus far been possible to measure indicate that this is true. On the other hand, it may be the combination of a proton and a closely packed electron. Such a condition would produce a particle of approximately but not quite the right mass that is electrically neutral, as the proton and electron would neutralize their charges. If this is the true nature of the neutron, it could account for one phenomenon constantly observed in the study of atomic physics. This is that beta particles, or electrons, are emitted from the nuclei of radioactive elements. There is no explanation of just where or how beta particles come from the nucleus except that in some manner they are produced from neutrons, since it is well established that electrons do not exist in the nucleus as separate entities.

Physicists formerly thought that electrons did exist in the nucleus, but it is now known that this cannot possibly be true. The fact that electrons as beta particles come out of the nuclei of radioactive atoms may be explained by assigning to the neutron this double personality. However, the difficulties in doing so are, first, to account for such a large particle as the electron being packed within the small volume of a neutron; and, second, the present determined mass of the neutron is not quite correct for it to be made up of the combined masses of a proton and an electron.

The reader should realize that any general discussion of such a fundamental question as the exact nature of the neutron is getting into the realm of the unknown. It is certainly true that conditions exist and happenings occur within the inner structure of nuclear particles that are little understood, and they may be entirely different from those in less compact spaces with which

we are familiar. Future research may give more information on this most interesting problem.

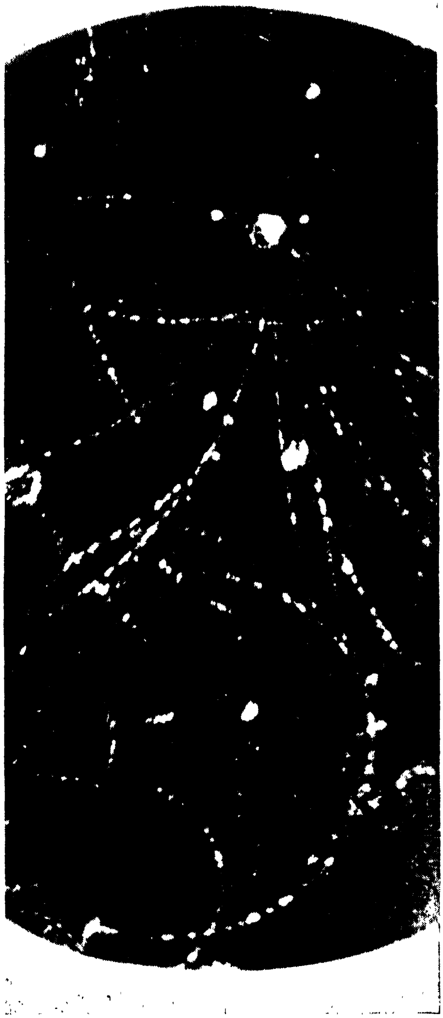
Positrons

During the years immediately preceding 1932 the fundamental particles making up the physical world had been universally supposed to be protons and electrons. Out of these two primary particles all the ninety-two chemical elements, and therefore all the thousands of compounds of the earth, were believed to be formed. From 1930 to 1933 the characteristics of the neutron were investigated, and the investigations indicated that the neutron was a particle of mass without electrical charge. Such indications were beginning to create doubt as to the correctness of the belief that all matter consisted of protons and electrons. Then, in August of 1932 Carl D. Anderson, while making a study of cosmic rays at the California Institute of Technology, took a photograph that served to upset still more the scientists' certainty regarding the structure of matter. He at once realized the significance of the photograph, and the report is that he and one of his fellow workers spent the entire night trying in vain to explain it without upsetting the prevailing idea regarding the structure of atoms. This, however, could not be done.

The photograph revealed the path of a free positive charge that had been ejected from the nucleus of an atom when it was bombarded by cosmic rays. Cosmic rays are high-energy radiations which come from the cosmos, or outer space. When they are absorbed by the nuclei of atoms, the nuclei are disrupted, often emitting the well-known alpha particles, as well as protons and electrons. However, this photograph gave evidence that a positive charge devoid of the mass of the proton had been ejected from the nucleus of an atom. The cosmic ray had evidently "kicked out" of the nucleus an identical twin, except in electric charge, of the free negative charge, or electron. It was the first time in man's scientific history that even a fleeting glance had been secured of a positive charge separated from the mass of the proton coming from an atom. The particle was called the positron.

Since this first discovery many similar results have shown that positrons are formed when atomic nuclei of a substance,

such as lead, are bombarded with high-energy particles or high-energy radiation. In many cases when a positron is formed, an electron is also produced. The two opposite electric charges are, of course, deflected in opposite directions when they are caused to move through a magnetic field, as is shown in the accompanying photograph of tracks of positrons and electrons passing through such a field. This opposite deflection of the particles was what led to Anderson's discovery of the positron. The length and curvature of the paths give definite information regarding the mass and energies of the ejected particles; and their curvature in the opposite direction shows that they possess opposite electric charges.



Tracks of particles emitted from atoms by bombardment with cosmic rays. Particles bent to the left are negative electrons; those to right are positive electrons (positrons). (Photograph by Arthur H. Compton.)

A study of many photographs taken of the paths of ejected positrons has given us considerable information regarding them. The positron has almost exactly the same mass as the electron, which is, it may be remembered, 1/1,835 that of the proton. Furthermore, the length of the paths reveals that free positrons have only a very short "life," usually of the order of a millionth of a second. Their demise is produced by union with a free elec-

tron, of which there are many in any space containing atoms, when both particles suffer the fate of complete annihilation, becoming instead radiant energy. Anderson, however, has said that the positron, if removed from a region of negative electrons, may live for millions of years instead of perishing in a millionth of a second. He believes this condition to exist in stellar space but not upon the earth.

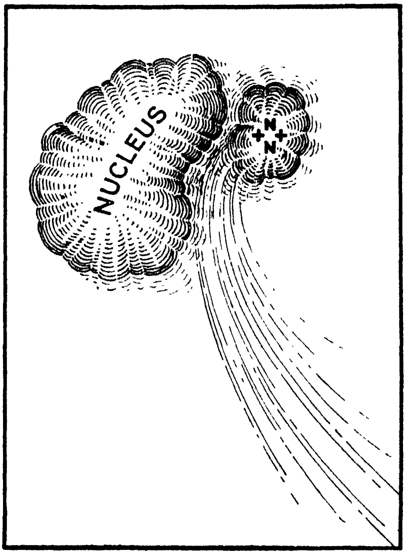
Anderson's discovery showed that the positive unit charge could exist entirely independent of the mass of the proton and that the proton with its relatively heavy mass is not the unit of positive electric charge or at least not the only unit. However, the general reader must not infer from this discussion that the proton is an insignificant unit in the structure of atoms, for just the reverse is true. It is always found within atomic nuclei. The formation of the positive charge separated from the mass of the proton occurs only rarely and under very special conditions in which the nucleus is bombarded with high-energy particles or radiation. Whether the positron is stripped off the proton or is itself created within the nucleus from the impinging energy is not known. What is known is that the energy goes in and the positron comes out.

The Nucleus as It Is Known at Present

So, when all is said, what is the simple picture of the structure of the nucleus of the atom? The answer is that the picture is not a simple one when it includes many details. However, if we may take the same liberty that an artist takes in painting, when he creates a general impression rather than an analysis of his theme, we may reduce the picture of the atomic nucleus to reasonable simplicity.

The latest (at the present writing) accepted theory of the structure of atomic nuclei is that they are compact entities consisting of protons and neutrons. Within the nucleus is concentrated practically all the mass of the atom, also all the positive charges. The size of the nucleus has been measured with considerable precision within recent years. It is interpreted as its "sphere of influence" rather than any definite wall or boundary, which means that an impinging particle may approach the vicinity without producing any material changes either in the

particle or in the nucleus. Somewhat the same idea may be given by imagining just how close one billiard ball may come to another without bouncing off it or moving the other ball. Measurement of this sphere of influence of the nucleus indicates that it is of the order of 10^{-12} to 10^{-13} centimeter in diameter. The diameter varies somewhat, depending upon the complexity of the nucleus as found in different chemical elements.



The size of an atom nucleus is interpreted as its "sphere of influence" within which an ordinary impinging particle does not penetrate.

The nuclei of the atoms of the different chemical elements are supposed, except for the hydrogen atom, to be different combinations of protons and neutrons. The hydrogen atom is the simplest in structure of all the elements. Its nucleus consists entirely of one proton, constituting one unit of weight and one positive nuclear charge. Helium atoms are the next simplest in structure, with the nuclei each having two protons and two neutrons, giving them an atomic weight of 4 and a positive nuclear charge of 2. The chemically active metal lithium has atoms with nuclei containing three protons and four neutrons, which give to each of them an atomic weight of 7 and a positive nuclear charge of 3. The nuclei of atoms of the other chemical elements are likewise built up of other combinations of protons and neutrons. The atom of oxygen, for example, has eight protons and eight neutrons in the nucleus, giving it an atomic weight of 16 and a positive nuclear charge of 8; gold has an atomic nucleus containing 79 protons and 118 neutrons, making its atomic weight 197 and its nuclear positive charge 79; and so on for the other chemical elements.

In the minute speck of matter that is the nucleus, the ordinary repulsion between protons and neutrons gives way to a newly discovered inner nuclear force which is very strongly

attractive. This so-called "binding energy" serves as a storehouse for very large quantities of nuclear energy. Therefore, great quantities of energy applied in high-speed atomic "bullets" are required to disrupt the nucleus of an atom. The repulsive forces between atomic particles serve as an enormous electrical barrier which can be overcome only with difficulty. This means that if one wishes to bombard an atomic nucleus with another atomic particle, it is necessary to speed these latter particles at the target with enormous amounts of energy.

However, when atomic nuclei do "explode" under bombardment with high-energy missiles, or naturally as in the case of radioactive atoms, great quantities of energy are released. Three well-known types of rays, the alpha, the beta, and the gamma, are emitted. Under special conditions a fourth ray, the positron, is also ejected. The alpha particles are helium nuclei consisting of two protons and two neutrons, as previously noted. Also, it should be remembered that beta rays are electrons. How the beta particle can be emitted from the atomic nucleus is not understood unless it is in some manner created from the neutron at the instant of ejection. It is almost equally difficult to account for the source of the positron. It seems to be a positive nuclear charge that has been removed from the proton or one that is actually created from high-energy radiation absorbed by the nucleus. The gamma rays are a form of radiant energy, similar in nature to X rays, which are released at the instant of atomic disruption.

Outside the Nucleus

The nucleus, or heart, of the atom is, we have seen, a real and definite thing; it is heavily laden and full of energy. However, in no sense does it constitute all the atom, for there is much outside the nucleus, certainly most of the volume and all the negative electric charges. The latter are more specifically called the extranuclear electrons, and the number contained in an atom is determined by the particular chemical element to which the atom belongs. The hydrogen atom, for example, has one such electron, which gives it a negative electric charge of one. This negative charge is exactly balanced by the one positive charge

of the proton in the nucleus, thus making the normal hydrogen atom electrically neutral. The helium atom contains two extranuclear electrons, these two negative charges being electrically neutralized by the two protons in the nucleus. Likewise, lithium has three outer electrons; oxygen has eight; gold has seventy-nine; and in each case the number is the same as the number of protons in the nucleus. The normal atoms of these elements are electrically neutral, therefore, and these same conditions apply to all the ninety-two chemical elements.

The structure and characteristics of the atom outside the nucleus may, in a scientific sense, be said to be understood; *i.e.*, the general properties of the extranuclear electrons are now well known. The mass and electric charge of the electron have been accurately measured, and the number of outer electrons surrounding the nucleus of the atom of each chemical element has been determined. Also, it is known that the extranuclear electrons are in a continuous state of fast motion such as to constitute a sort of aura, or cloud, of relatively tenuous nature around the nucleus.

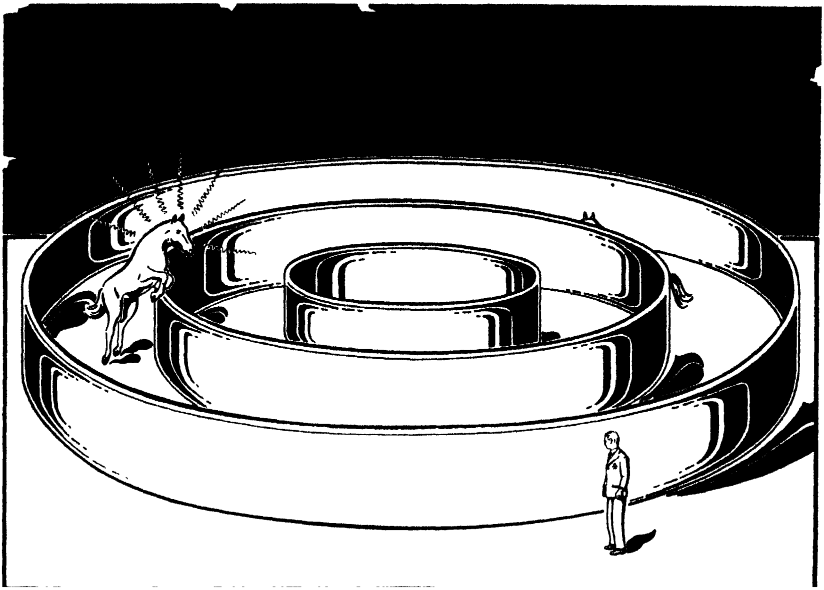
We are somewhat at a loss in attempting to present a graphic picture of just how this motion is taking place. It is understood and explained in terms of mathematical equations but not in terms of any mechanical models. For example, the electrons are not moving around the nucleus in fixed paths or orbits comparable to the fixed orbits of the planets around the sun. Furthermore, they manifest characteristics that indicate that they are not rigidly shaped particles, as the earth is a particle with a definite size and shape in the solar system; rather they have a tenuous, wavelike form. Since it is, of course, impossible, because of their minute size and excessive speeds, actually to see the electrons in the atom, their effects must be observed in order to learn about them. This will be much the same as observing the effects of machine-gun fire in order to learn something about the bullets that go whizzing through the air.

By man's use of indirect observations and brilliant mathematics the electrons have been made to reveal some of their secrets. They flicker around the nucleus in cloudlike traces. Although we know the exact probability of where an electron will be, this is the most that can be known as to its position at a

given instant. We know that it moves around the nucleus with velocities of some 10^8 centimeters per second, which is about 630 miles per second. (The blades of a revolving fan are slow in comparison.) This velocity gives the electrons great momentum, or "energy of motion."

One other exceedingly important characteristic of extranuclear electron movement has been discovered, namely, that as the electrons speed around the nucleus they may be transferred to positions nearer to the nucleus or to other positions farther away. To state this more accurately, they may move around the nucleus in levels of different energy states. For the sake of simplicity, these different energy states may be compared to the different floors in an apartment building or residence hall. When a person is on the lower floor of the building, he has zero potential energy with respect to the ground. When he is on the second floor, he has more potential energy; and when he is on an upper floor, he has still more. In its simplest terms this means that the potential energy of a person on the upper floors would cause him to fall to the ground were the floors removed, but at the ground level he possesses no such energy. Likewise, an electron may occupy a "ground" energy state or a higher energy state. While in a higher energy state, it has more energy of motion (which it received from an outside source) than it had while in the ground energy state. As a matter of fact, an electron may occupy within the atom a large number of different energy states. They may be thought of as concentric spheres with the nucleus of the atom approximately at the center.

With certain exceptions, so long as an electron continues its motion in one energy state, no energy is absorbed or emitted by the atom. This condition has been called the "stationary state" of the atom; and it is analogous to saying that so long as a person moves around in an apartment on one floor in a building, no potential energy is acquired or used up. This same thing may be visualized by imagining a horse in a pasture field that is fenced off into concentric circles by high board fences. While the horse runs in one circle, he cannot be seen by an observer on the ground. Only when he jumps over a fence into another circle does his potential energy change; only then can he be seen. At that time only does he "give out" visible radiation.



A horse moving in these concentric pastures would "emit" visible radiation so as to be seen by an observer on the ground only when he jumped from one field into another.

When an electron jumps from one energy state to another, energy is either absorbed or radiated, just as when a person goes from one floor to another he will gain or lose potential energy, depending upon whether he goes up to a higher apartment or down to a lower one. Now, suppose that an electron moves from a lower to a higher energy state; it must secure more energy to do so. In this case it absorbs energy somewhat as a person must acquire potential energy in going from a lower floor to a higher one. The absorbed energy in the case of an electron must come from an external source. On the other hand, when an electron moves from a higher energy state to a lower, energy is emitted in the form of radiation, just as a person loses potential energy in descending a flight of stairs.

Even as an electron moves around in a given energy state, it does not follow a definite path; rather, this path is spread out over the whole space of the energy state in a sort of cloud. The most that can be done is to calculate the probability of where the electron will be in the cloud at a given instant. The density of

the cloud shows therefore, the probability of finding the electron at that point; and electrons found at different energy states would, of course, produce different clouds. Using somewhat this sort of reasoning, it is possible to construct models of an atom to represent the electrons in different energy states by visualizing concentric clouds of different densities to represent different energy states. Photographing such models provides the pictures sometimes used to represent different energy states.

Composite Summary

In a general way, then, a composite view of the structure of the atom may be had. The basic plan of its architecture is made up of two distinct divisions, the nucleus and the extranuclear electrons. The former is composed of protons and neutrons, extremely minute in size. Each of these particles contains one unit of mass, and the protons each have in addition one unit of positive electric charge. The charge may be the positron bound intimately to the unit of mass. This, however, is conjecture rather than fact, and for all practical purposes the proton is to be considered at present as the unit positive charge with one unit of mass. The nuclei of most atoms are extremely stable, being held together with enormous forces; but when they are disrupted, various particles come flying out, and large quantities of energy are released.

Outside the nucleus are one or more electrons, or unit negative electric charges. Each of them has only 1/1,835 of the mass of a proton. These electrons are moving at high speed in a number of different ways, and they occupy different energy levels under certain conditions, as mentioned above. Electrons are the less stable parts of atoms. Their speed and energy states may be easily changed; some of them may even be stripped temporarily from the atom itself by a chemical change or merely by the substances being dissolved in water.

Alchemist's Dream Realized

Through the centuries man has cherished many ideas that have seemed at times to be fantasies. He hoped to find a universal solvent, a philosopher's stone, and to bring about the change of one element into another or the transmutation of the

elements. Chemistry and physics alike had their ancient origins in the mystical efforts of alchemists to achieve these ends. Out of twentieth century research into the atom has come the realization of the dream of transmutation, and even the long-sought transmutation of other elements into gold has been reported. The quantities thus achieved, however, have been minute, and, since they promise to remain so, there is no possibility of disrupting the gold standards of the world. The significance of transmutation lies in giving us a wider understanding of the nature of atomic structure.

Natural transmutation of the elements occurs in all radioactive substances and has been known since Becquerel's discoveries. It was Lord Rutherford, in 1919, who first achieved artificial transmutation when he bombarded the atoms of nitrogen gas with alpha particles which came from radium and obtained oxygen and hydrogen gases as a result. The technique of this experimentation was based upon a knowledge and understanding of the structure of atoms; *i.e.*, the fundamental difference between the different elements is known to be in the make-up of the nuclei of the atoms. In order to change one element into another it is necessary to change the nucleus of its atom. Rutherford reasoned that this could be done only by shooting into the nucleus a particle small enough to enter and containing enough energy to disrupt the stable conditions existing there. To change each atom would require, of course, an individual hit on the nucleus.

The explanation of transmutation may be understood if one consults the Periodic Table of the elements (such as the one shown on page 181) and notices what may happen with a change in nuclear structure. Let us see how this works out in the case of bombarding nitrogen with an alpha particle.

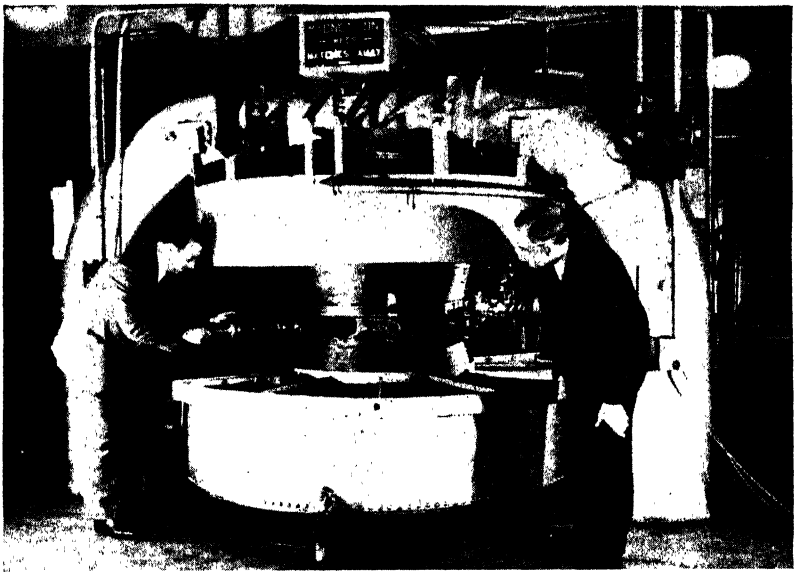
The nitrogen atom has an atomic weight of 14 and a nuclear positive charge of 7, which means that the nucleus consists of 7 protons and 7 neutrons. When an alpha particle containing 2 protons and 2 neutrons succeeds in entering the nucleus, the nitrogen nucleus then contains 9 protons and 9 neutrons, giving it an atomic weight of 18 and a positive charge of 9. However, this combination is unstable and immediately breaks up, giving off one proton. The proton is, of course, the nucleus of the

hydrogen atom. It soon picks up an electron from its surroundings and becomes a normal hydrogen atom. The particle remaining after the emission of the proton contains 8 protons and 9 neutrons, which gives it an atomic weight of 17 and a positive charge of 8. By consulting the atomic chart, it is seen that the only element having a positive nuclear charge of 8 is oxygen. The atomic weight of our particles is 17, however, rather than 16; but this additional weight is due to the presence of an additional neutron. Neutrons do not influence the chemical properties of an element, and so the remaining particle is a "heavy" oxygen atom, or what is more specifically called an isotope of oxygen. We have, therefore, obtained the elements hydrogen and oxygen by bombarding nitrogen with alpha particles.

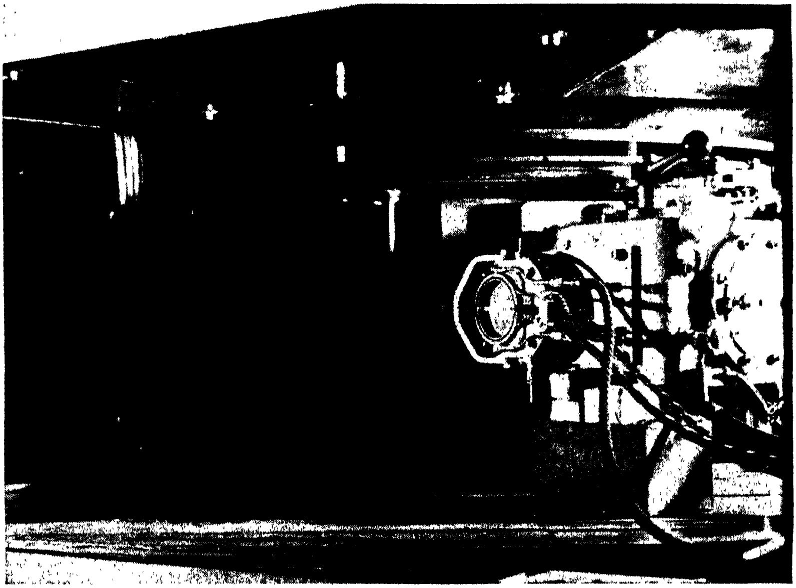
Rutherford had to work with the relatively simple techniques known at that time. A speck of radium provided the alpha particles. The only way to observe their effect was to watch through a microscope the tiny flashes of light produced when one of these alpha particles struck a fluorescent material. Seated in a dark room, Rutherford tediously counted a few faint scintillations originating from the bombardment of the nitrogen by the alpha particles. Chemical tests then revealed that a few of his nitrogen atoms had been disrupted and changed into oxygen and hydrogen, thus proving chemically the theoretical deductions and experimental observations. The process was laborious, and the results obtained infinitesimal; but the significant fact was that it had been accomplished.

Immediately following this transmutation, other investigators took up the work. Rutherford's experiments were verified; a number of other elements were likewise bombarded with alpha particles; and in some cases transmutations were secured. However, radium, the source of these particles, was expensive and scarce. The supply of atomic "ammunition" was limited, thus delaying wide experimentation in this field for a time. Then in 1932, with the discovery of the neutron and also of "heavy" hydrogen by Prof. Harold O. Urey of Columbia University, a new source of atomic particles suitable for bombardment became available.

Although the early work on transmutation was achieved with the "bullets" emitted from the natural radioactive ele-



Operation of cyclotron at Columbia University being explained by Dr. John R. Dunning who is adjusting exit chamber of high-energy particles. (Photograph by courtesy of J. R. Dunning.)



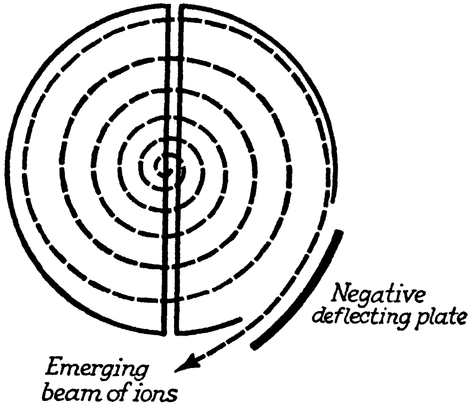
A beam of light from ionized air produced by high-energy particles from exit chamber of cyclotron at right. (Life Magazine photograph of cyclotron at University of California.)

ments like radium, scientific research soon disclosed that it was possible to build machines that would speed atomic particles, such as neutrons or heavy hydrogen, with even greater energies than those obtained from the natural radioactive materials. These machines are of several types, of which the most fruitful has been the cyclotron, invented by Prof. E. O. Lawrence of the University of California. When natural radioactive elements disintegrate, the particles coming out of them are emitted with energies of about five million electron volts; but with the machines that have been built, man has been able to impart to atomic particles far greater energies. The cyclotrons already in use produce particles speeded with energies equivalent to thirty-two million electron volts. An electron volt, it should be explained, is only the scientist's unit of energy in the realm of atoms. Professor Lawrence has just received a grant of $1,250,000 for the fabrication of an enormous cyclotron which will have a total weight of nearly five thousand tons and will produce particles having energies greater than one hundred million electron volts. Such great energy exceeds by far the energies of particles from radium or other radioactive materials and approaches that found in nature only in the mysterious cosmic rays.

Professor Lawrence's cyclotron accelerates the atomic particles by an ingenious electrical method which in many ways resembles the slingshot used by David in his conquest of Goliath in the old Biblical story. The electrified atomic "bullets" are made to travel round and round in a pancake-shaped device which is inserted edgewise between the poles of a large electromagnet. Ordinarily such an intense magnetic field would make the atomic particles travel in circles; however, twice during each revolution they are speeded up in small steps by an electrical potential. As a result they travel in a spiral path of increasing radius within the cyclotron. After several hundred revolutions, which take only a very brief instant of time because of their high velocities, the particles have built up their maximum energy and pass through an exit chamber, where they are used by the scientists to bombard any elements with which they wish to experiment. The cyclotron contains complex electric devices for counting automatically the number of hits and trans-

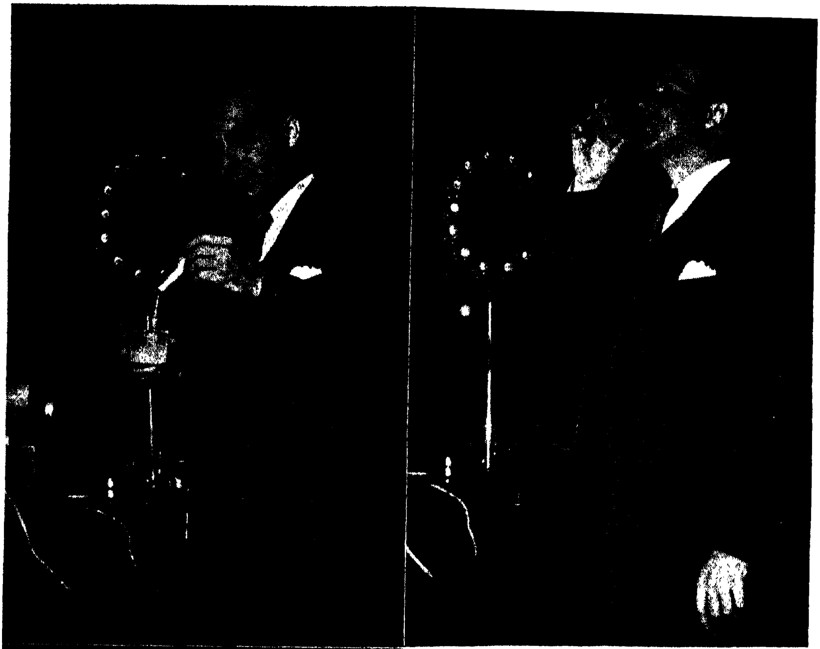
formations made in such bombardments and automatic cameras for photographing the paths of particles coming out of the bombarded nuclei so that records of millions of bombardments may be made.

The most common type of experiment performed with cyclotrons is the bombardment of an element like copper, for example, to produce in it an artificial radioactivity. The activity is then detected and measured by suitable apparatus. Artificial radioactivity results from the copper's having been changed to another atom that is unstable and that breaks down, emitting particles or radiant energy. Transmutation has been produced in nearly all the chemical elements now known. More than 200 radioactive varieties, or isotopes, of these chemical elements have now been created. All of them spontaneously disintegrate in a manner quite similar to that of radium. Among the most striking of the radioactive isotopes is that of sodium. It is produced simply by taking ordinary table salt and bombarding it. The radioactive salt liberates gamma radiation quite like that from radium and has been suggested as of therapeutic value in the treatment of cancer.



An electrified atomic particle admitted into the center of the two-section chamber of the cyclotron is caused to travel in a spiral path with increasing velocity by a strong alternating magnetic field, and is finally directed out of the cyclotron by a deflecting plate.

There are some striking differences between such nuclear interactions and ordinary chemical reactions. One is that for each atom taking part in a nuclear change a much greater amount of energy is involved than in a chemical reaction, usually many millions of times more than in the most violent chemical reactions, such as the explosion of dynamite. Another difference is in the small number of atoms that undergo nuclear change; therefore, the total amount of the new element that can be produced is extremely small. At present rates of artificial



In these three photographs Dr. Dunning is shown, first, dissolving some table salt that has been made radioactive by bombardment with high-energy neutrons from a cyclotron; second, drinking the solution; and third, using the radioactivity after the salt has been dissolved in the blood stream and circulated to his hand to actuate a Geiger amplifier and thereby light the lamps. (Black Star photograph.)

atomic disintegration all the most efficient apparatus in this country operating continuously for a hundred years could produce less than a penny's worth of helium, and the power cost alone would approach a million dollars.

At the present writing, atom transmutation is not accomplished on a commercial scale; however, a very extensive program of research in the production and study of it is in progress. In 1939 a discovery of primary significance regarding atomic disintegration was made, one that has been investigated on a wide scale during 1940 and that now seems to offer a possibility of utilizing the enormous energy of heavy atoms. The heavy atoms of uranium, when bombarded with low-energy neutrons, were found to break down into two lighter atoms, each of which undergoes further disintegration with the liberation of great quantities of energy. Apparently the heavy uranium atom lies close to the limit of complete instability, and even a low-energy impinging particle may cause this limit to be exceeded and the atom to fly apart. Such a process has been compared to that in which a large drop of liquid splits easily into two drops when set into proper vibration. This unusual process of splitting the uranium and other heavy atoms into two parts that emit large amounts of energy has been called fission.

When the uranium atom is bombarded with a low-energy neutron, it apparently splits into an atom of xenon and an atom of strontium. The xenon then goes through a series of disintegrations and finally becomes stable cerium. The strontium is likewise unstable, and it, too, undergoes successive disintegrations to produce stable rubidium. In the process of these changes, a total of approximately two hundred million electron volts of energy is liberated. Some idea of the significance of these energies may be obtained when it is realized that the energy of the alpha particles from the naturally radioactive radium is in the order of five million electron volts and that the energy of the impinging neutron necessary to split the uranium is far less than this.

The photograph at the beginning of this chapter is one that gives a visible relative measure of the enormous energies liberated in fission. The high energies resulting from the disintegration of xenon and strontium were used to actuate a cathode-ray tube, the front end of which is shown in the picture. The small

dots along the base line represent energies of approximately the same value as those of alpha particles from radium; and the long vertical lines, those produced by fission of a uranium atom.

Furthermore, it has been found that when fission occurs in the heavy atoms, not only are these high energies liberated, but also low-energy neutrons are emitted. The neutrons in turn have been found to produce fission in still other uranium atoms in a sort of chain fashion so that the action repeats itself. Should a practical method be found to cause this chain action to continue indefinitely, an extremely cheap production of an enormous amount of energy might be procured. There is at present no indication of how to make the chain action continuous over a period of time, but the possibility of such discovery does exist.

Aside from the possibility of any such cheap source of power, the artificial transmutation of the elements has given man one of his greatest techniques of research. Such research has yielded us precious information on the structure of all atoms and seems likely to go a long way in solving the secrets of the true constitution of the material universe.

REFERENCES FOR MORE EXTENDED READING

HARRISON, GEORGE R.: "Atoms in Action," William Morrow & Co., Inc., New York, 1939.

It is somewhat unusual for a distinguished scientist to write a popularized book in his own field. Professor Harrison has presented in an interesting and fast-moving style the contributions of the science of physics, in terms of atomic forces and properties, to everyday living and illustrated his book with a number of excellent photographs.

RUSK, RODGERS D.: "Atoms, Men and Stars," Alfred A. Knopf, Inc., New York, 1937.

A popularized discussion which gives, in Chaps. VII, VIII, IX, a general concept of atomic structure and also points out certain aspects of the significance and meaning of recent discoveries in this field.

BAZZONI, CHARLES B.: "Energy and Matter," The University Society, New York, 1932.

This little book presents a summary of knowledge relating to the structure of matter at the time of its publication. It is written for the layman in a style that uses a minimum of technical language.

Hull, Gordon F.: "An Elementary Survey of Modern Physics," The Macmillan Company, New York, 1936.

This text is written for advanced college students who may be majoring in English, history, or another unscientific subject and wish to become familiar with the general phases of "modern physics," which is primarily atomic physics. Chapters 2, 3, 6, 11, 14 give a good account of atomic structure without presenting serious difficulties in mathematics to the superior general student.

Smyth, H. D., and C. W. Ufford: "Matter, Motion and Electricity," McGraw-Hill Book Company, Inc., New York, 1939, Chaps. XIX, XX, XXI, XXVII.

The chapters referred to include a discussion of atomic structure, the properties of moving atomic particles, and an abbreviated description of experimental techniques in atomic physics. The material is suitable for superior college students with a general understanding of physics.

Millikan, Robert A.: "Electrons (+ and −)," University of Chicago Press, Chicago, 1935.

A thorough and comprehensive account of atomic structure as understood in 1934, presented in a lucid style that is relatively free from mathematics. Recommended for the superior student who wishes a detailed understanding of the major discoveries in atomic physics during the first third of the twentieth century.

Speakman, J. C.: "Modern Atomic Theory," Longmans, Green & Company, New York, 1938.

This book is an elementary introduction to the study of the atomic structure of matter and is suitable for the intelligent reader with little previous training in physics or chemistry.

Darrow, Karl K.: "Introduction to Contemporary Physics," D. Van Nostrand Company, New York, 1939.

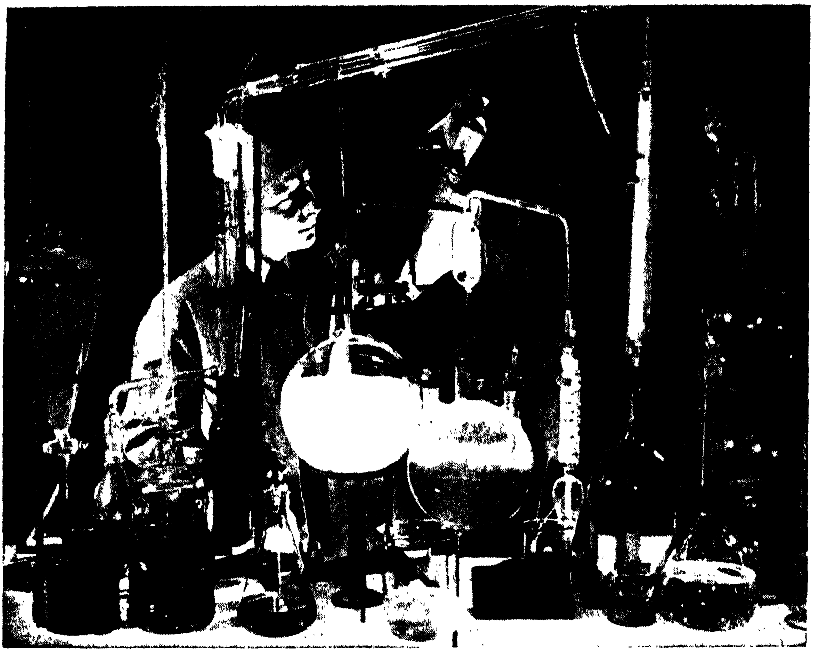
A thorough discussion of nuclear physics, electron diffraction, neutrons, and artificial radioactivity, with a minimum of highly technical material, by a competent physicist and author.

The Yale Scientific Magazine, published quarterly by the Yale Scientific Magazine, New Haven, Conn.

Popularized articles on scientific topics of timely interest to students and laymen; also a digest of the latest science news.

The Physical Review, published by The American Institute of Physics, New York.

A monthly journal devoted to articles on research in the field of physics, many of them relating to atomic structure.



U. S. Department of Agriculture (Peter Kellian).

5: REACTION

Whereby Chemical Elements Form Compounds

IN THE Dictionary of Applied Chemistry and the Dictionary of Organic Chemistry we are told that some 300,000 different compounds of carbon are now known to man. Each one is a different kind of substance. Carbon compounds include many of the substances and tissues of the human body and of the bodies of other animals and plants as well as a host of other materials. These same sources reveal that there are many thousands of additional compounds of other elements, these compounds constituting such materials as ores, rocks, precious stones, and soil. Altogether nearly a half million different kinds of substances are now known to be on the earth! Yet this multitude is formed from just ninety-two different chemical elements.

These compounds are not all by any means in a fixed and static condition. Many of them are ever changing. Even while one is reading this sentence, many thousands of chemical changes are taking place in his body. The rusting of every piece of iron, the burning of every fire, the growing of every blade of grass is chemical change. These processes are everywhere about us. An understanding of the atomic structure of matter gives us some inkling of how such a variety of things are provided from the relatively few fundamental elements. All the materials referred to above are known to consist of combinations in various ways of atoms of the ninety-two different chemical elements, and the chemical changes going on in and about us are manifestations of reactions between such atoms.

Units of Chemical Changes

If we knew that the destiny of all mankind was controlled by just ninety-two people, it would behoove us to know something about these individuals. The fact that ninety-two chemical elements compose all the materials of the earth indicates the importance of each. What, then, is a chemical element?

Suppose that a piece of gold is taken into the laboratory and cut into two pieces; both of them will still remain gold. These pieces may in turn be divided in halves, and the process continued indefinitely, and there would still be no actual destruction of the gold. The original gold piece might be heated to extreme temperature until it was melted or possibly even turned into a vapor, yet it would remain gold. Should it be treated by passing an electric current through it, it would not yield; in addition it cannot be reduced to a simpler substance by chemical action. Furthermore, it has been found that all its atoms are essentially alike; it is composed of only one kind of matter. Gold is therefore elementary and is called an element.

One illustration showing a few of the chemical properties of an element may serve to give a little better conception of the nature of an elementary substance. When a piece of yellow sulphur is put into a container and heated, it will melt to form a viscous, dark-colored liquid; if left to cool, it will again solidify to the familiar yellow material that it was in the beginning, indicating that no change in its chemical composition has

Pouring liquid gold into molds at a U. S. mint for shipment to the depository at Fort Knox, Kentucky. (Wide World photograph.)

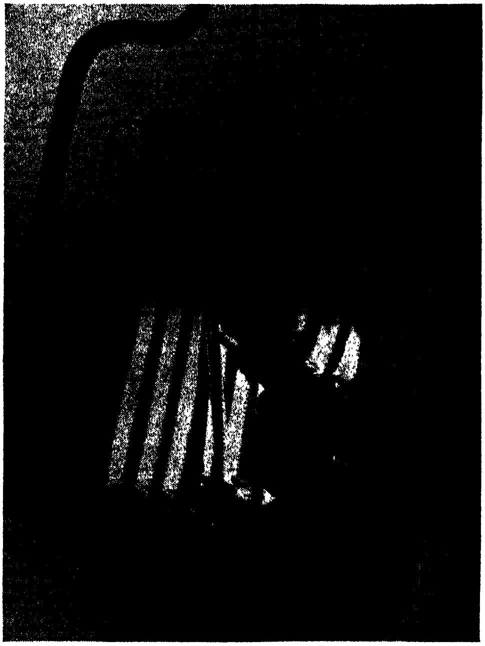
occurred. However, should a stream of hydrogen gas be bubbled through the liquid sulphur, the sulphur will eventually disappear. During the process an invisible gas which has the strong odor that we associate with spoiled eggs is formed. Obviously, something has happened to the sulphur, and other experiments have shown that it combined with the hydrogen to form the pungent gas hydrogen sulphide. Let us now collect some of this gas, heat it in the proper kind of burner with a shortage of oxygen, and hold a cold porcelain dish over the flame. A layer of yellow sulphur crystals is deposited on the cold dish. A second reaction has occurred in which the hydrogen sulphide has been decomposed and the sulphur has been recovered in its original form. The sulphur has gone through two chemical reactions, but no change has occurred in its fundamental nature.

Sulphur is an element, and, likewise, all other substances that show these elementary properties are elements. The general statement may be made, then, that an element is a substance that resists all ordinary or chemical efforts to decompose it into simpler substances; also, it may enter into chemical reactions without alteration of its fundamental nature.

Probably gold, copper, tin, and iron were the four elementary substances first recognized as such by mankind. At the beginning of the Christian era the scientific world knew of only seven elements—the four just mentioned and tin, lead, and mercury. During the next 1,700 years only ten more were added to the list. Then in 1774 an English clergyman and chemist, Joseph Priestley, and the distinguished Swedish chemist, Karl Sheele, independently separated pure oxygen from one of its compounds and discovered that oxygen was an element. Since that time a total of ninety elementary substances have been identified and studied, and two others have been reported discovered, making ninety-two in all. This probably constitutes the entire number of elements existing in the earth.

Some of the elements are very common on the earth; others are scarce. Oxygen, for example, constitutes one-half the known crust of the earth, and silicon makes up another fourth. The next most abundant element is aluminum, which forms about 7 per cent. Iron comes next, and it accounts for about 5 per cent of the known earth's crust. Remarkably enough, carbon constitutes less than one-tenth of one per cent, even though it is a constituent part of all living things, and its compounds are so numerous that they greatly exceed in number those of all the other elements combined. Many of the elements have never been seen in pure form by most chemists, so rare are they. In fact, some seventy-five of the elements make up less than one per cent of the earth's crust, and at present they are mere curiosities with no known use. However, the future may discover important commercial or medical use for some of these little known and rare elements. Less than half a century has elapsed since Madame Curie discovered radium and learned of the properties that make it so valuable today. Aluminum metal was a laboratory curiosity with a monetary value of $100 a pound until Charles M. Hall discovered in 1886 a practical method of liberating it from its binding compounds.

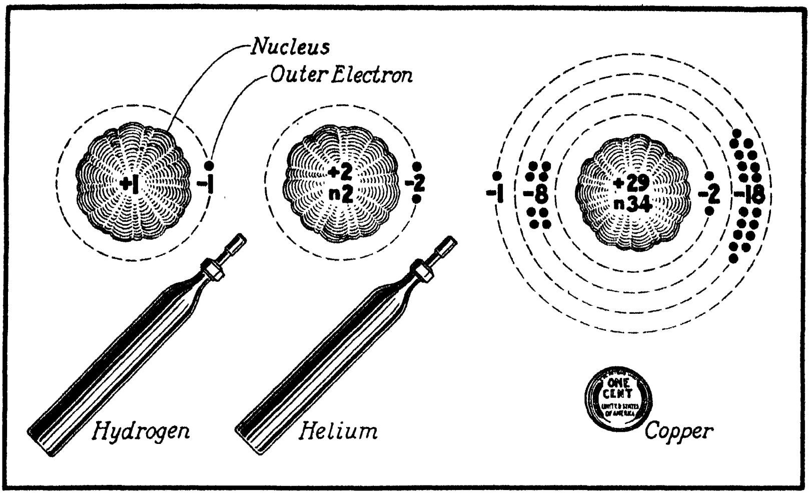
Elements are identified by certain characteristics or properties that set them apart from other substances. These are such things as their color, hardness, odor, taste, density, solubility in water or other liquids, melting point, boiling point, and ability to conduct electricity. The ability to burn in air and the



The magnified image of a tungsten filament in a lighted lamp is projected on a screen so that the filament's properties of expansion and buckling may be accurately measured. (Courtesy of Westinghouse.)

manner in which an element enters into chemical reactions with other substances are also properties that aid in distinguishing one element from another. For example, oxygen is a colorless, odorless, tasteless gas with a density at standard conditions of pressure and temperature of about one seven-hundredth that of water. It is slightly soluble in water, and it liquefies at about 180° below zero on the centigrade thermometer. It readily combines with many elements to form oxides, such as carbon dioxide and water vapor (hydrogen oxide), as well as many other compounds.

Likewise, these properties determine whether a substance is useful industrially or commercially. Copper would not be of such great value if it were not an excellent conductor of electricity. Carbon and hydrogen are useful partly because they combine readily with oxygen, *i.e.*, burn, and in so doing liberate great quantities of heat. Tungsten is valuable because of several remarkable properties. It is a hard metal which does not melt



Graphic representation of three chemical elements.

when heated to temperatures high enough to produce a white light. It does not rust or tarnish and is not attacked by cold acids. Thus it is exceedingly useful for electric-lamp filaments and for high-speed cutting tools. As such, it is now considered one of the key metals of industry.

The nature of the ninety-two different elements may be simplified somewhat by recalling some of the essential features of atomic structure. The ultimate particles of all the different atoms are neutrons, protons, and electrons; therefore, the only way in which these ninety-two chemical elements differ from each other in their composition is in the number and arrangement of these particles in their atoms. Likewise, many of the properties of the elements, both physical and chemical, are accounted for by the number and arrangement of these particles, particularly the extranuclear electrons.

To review briefly, we recall that hydrogen is the simplest element, consisting of a nucleus of one proton and one extranuclear electron. No simpler arrangement could exist in an atom. The next simplest element is helium, with a nucleus of 2 protons and 2 neutrons and 2 electrons outside the nucleus. Lithium is the lightest of all metallic elements, with a nucleus consisting of 3 protons and 4 neutrons, and 3 extranuclear electrons. The atom of sodium, the active metallic element forming a

part of table salt, consists of a nucleus of 11 protons, 12 neutrons, and 11 outer electrons. Copper has 29 protons and 34 neutrons in the nucleus and 29 extranuclear electrons. Uranium is the most complicated element known; its atom consists of a complicated nucleus of 92 protons, at least 142 neutrons, and 92 outer electrons.

The existence of 92 outer electrons in the uranium atom was the reason for originally assuming that there were at least 92 different elements, one for each numeral between 1 (hydrogen) and 92 (uranium), inclusive. All these 92 elements have now been discovered, with the exception of two very rare and elusive ones, and the latter have been reported found. No element with more than 92 outer electrons has ever been discovered, and it is likely that none exists. If additional elements did exist the atoms would be so complex that they probably would break down by atomic disintegration. In fact that is what is happening to uranium, as it is one of the radioactive elements.

Now, the hydrogen atom has one negative charge, corresponding to its one extranuclear electron. The helium atom has two negative charges, corresponding to the two outer electrons. Lithium has three such negative charges; sodium has 11; copper has 29; and uranium has 92 extranuclear charges. This number, corresponding to the extranuclear electrons within an atom, is known as the atomic number. The extranuclear charge, or atomic number, is of great significance in the behavior of the elements, because it determines their chemical properties and is the key to most chemical reactions.

Products of Chemical Changes

Various combinations of the 92 elements produce all the forms of matter common to the earth; water, wood, sugar, gasoline, coal, sapphire, alcohol, and hundreds of thousands of others. These are chemical compounds. In contrast to elementary substances, a compound is produced by the chemical union of two or more atoms.

Probably the most important as well as the most familiar of all substances is water. Yet even pure water is not an element, as it may be separated into two entirely different substances. Water is, therefore, an example of a compound. Table salt is a

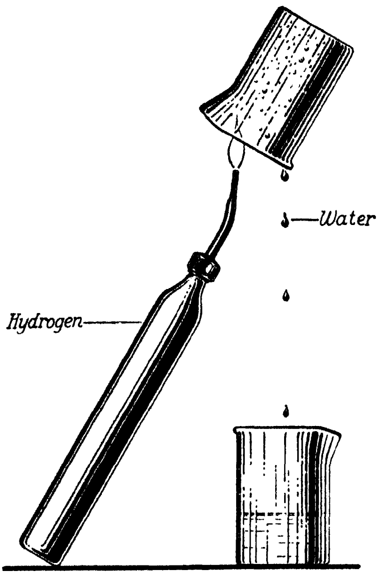
common and familiar material, and it, too, may be separated into other substances, showing it to be a compound. Water is produced by a union of hydrogen and oxygen, in the proportion of two atoms of hydrogen and one atom of oxygen. Salt is a compound of the elements sodium and chlorine, one atom of sodium uniting with one of chlorine.

The general idea of chemical compounds is that the atoms are more or less rigidly linked into groups called molecules. A molecule, then, is the smallest individual unit of a compound, and it will always consist of two or more atoms. In the sulphur demonstrations referred to earlier in this chapter, the molecules of the hydrogen sulphide gas formed contain one atom of sulphur and two of hydrogen. Anyone wishing to verify the formation of a compound from two elements may take a piece of sulphur and burn it in the air. When heated to its ignition point, it burns with a blue flame and is slowly used up. Although no visible product results, all doubts of a compound being formed from the combustion are dispelled with one breath of the choking gas sulphur dioxide that comes from the flame. In this case one atom of sulphur has combined with two atoms of oxygen in the air to produce one molecule of sulphur dioxide gas.

When elements enter into chemical combinations, their properties undergo a complete change. For example, sodium is chemically a very active metal, and chlorine is an active and deadly gas. When the atoms of these two elements combine, they form a mineral substance called salt. This is a crystalline solid, a substance necessary in everyone's food, and it is chemically neutral. As an additional illustration, hydrogen and oxygen are both gases that combine chemically to form water, a relatively neutral liquid. The chemical product of most of the explosion of the great hydrogen-filled Zeppelin *Hindenburg* at Lakehurst, N. J., in 1937 was water vapor.

The composition of a compound may be shown by either one of two methods, known as synthesis and analysis. Synthesis consists of putting together the elements to form the compound; analysis is the breaking down of the compounds into their constituent elements. When hydrogen and oxygen gases are mixed and heated to ignition by an electric spark or flame, they combine explosively, forming water vapor, and the gases disappear.

An accurate measurement of the combining gases shows that exactly twice as much hydrogen as oxygen by volume has been used up; and the compound, water, thus formed, is seen to consist of two parts hydrogen and one part oxygen. This is chemical synthesis. Then, again, the reverse of this process may be carried out. Water may be broken apart or decomposed by an electric current. This is accomplished by placing it in a glass U tube, known as the electrolysis apparatus, and passing an electric current through it. Two gases begin forming, and eventually the water disappears, leaving only the gases. One of the gases will have twice the volume of the other; and when they are tested, the one of larger volume proves to be hydrogen, the other oxygen. This is an illustration of chemical analysis.

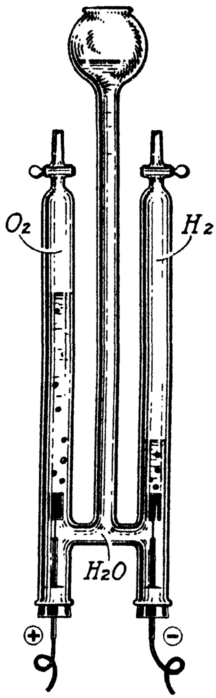


Synthesis of water.

We have seen that the very common substance water may be broken down into its constituent elements and also that these may be recombined to form water. By imposing different conditions upon the substances, the chemical reaction may be made to go in either direction. The double process of forming a given compound and then having it go back to the original substances is known as a reversible reaction. Under proper conditions reversible reactions may be brought about in all chemical changes, and they give us insight into some of the fundamental conditions of nature.

A significant fact regarding the energy involved in chemical reactions is shown in connection with reactions that are easily reversible. When oxygen and hydrogen combine, they do so with the liberation of a large amount of heat energy, as may be attested to by anyone who has ever observed the intensity of a hydrogen flame; *i.e.*, energy is always given off when oxygen and

hydrogen combine to form water. On the other hand, in order to separate water into hydrogen and oxygen it is necessary to add energy to the reaction. One convenient method of doing this is to add electrical energy, as in the electrolysis of water mentioned above. Herein lies the important point: The same amount of energy is required to separate a given quantity of water into its constituent elements of hydrogen and oxygen as is liberated by the same amount of those elements when they combine to form water. There is always an energy change with a chemical reaction. It may be that energy in some form is liberated, or it may be that energy is absorbed; but when the reaction is a reversible one, the two quantities of energy will equal each other, and a conservation of energy is always maintained.



Analysis of water.

Likewise, there is a conservation of mass in chemical reactions. A careful determination of the masses involved in the synthesis of water shows that the mass of water produced is exactly the same as the sum of the masses of the two gases uniting. Furthermore, in the analysis of water the combined mass of the hydrogen and oxygen formed will be precisely the same as the mass of water used up in the reaction. These same conditions hold true for the products of all other chemical reactions. Mass is neither created nor destroyed in such process. The conservation of mass is one of the laws of nature.

Oxidation, the Great Energy-liberating Process

The formation of water from hydrogen and oxygen is an example of an extremely important group of chemical reactions, called oxidation. These reactions are usually, although not always, the combinations of various elements with oxygen, and the oxygen may be the oxygen of the air. About 20 per cent of the air is oxygen, and the remaining 80 per cent is mostly nitrogen; however, oxygen is the most active element present in

Remarkable and intimate picture of the rapid oxidation of hydrogen gas in the Zeppelin Hindenburg, May 6, 1937. (International News photograph.)

the atmosphere, since it combines with nearly all other elements. The compounds thus formed are known as oxides. The fire in a gas burner is a good illustration of rapid oxidation, and so is the explosion of gasoline in the engine of an automobile. In fact all ordinary fires are due to the rapid oxidation of fuels, such as wood, coal, gas, tallow, alcohol, and kerosene.

When a match is lighted, it is quite evident that a chemical change is taking place, since the wood disappears with the liberation of heat and light, and only an ash remains to be seen. When the burning match is placed in an atmosphere containing only nitrogen, it immediately goes out, indicating that the chemical reaction involves a combination with oxygen. Such rapid oxidation is spoken of as combustion, or burning. The wood of the match is composed essentially of carbon and hydrogen. Most other fuels, such as the illuminating gas and gasoline mentioned above, are likewise compounds of these two elements.

When they burn, the carbon and hydrogen combine rapidly with the oxygen of the air to form carbon dioxide and water vapor, with the liberation of a great amount of heat energy. The reactions may be expressed in the following manner:

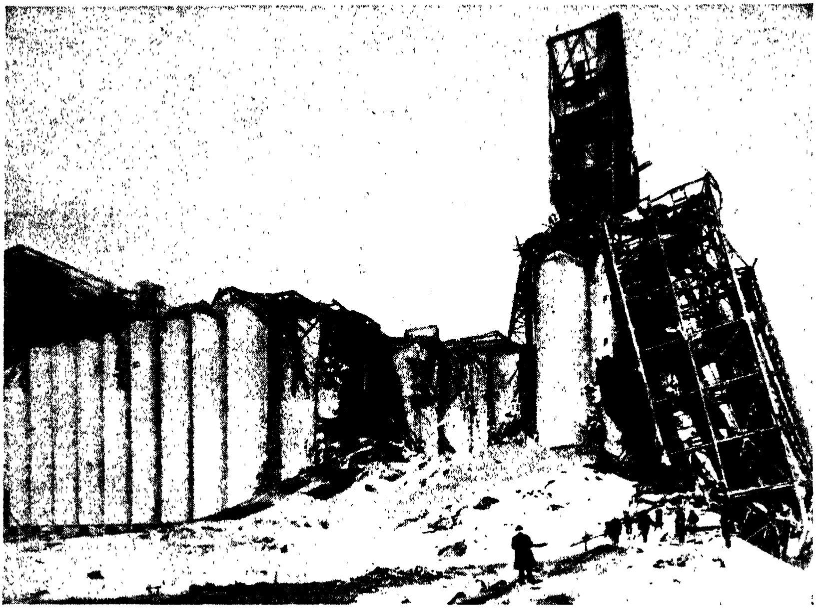
Carbon + oxygen → carbon dioxide + heat
Hydrogen + oxygen → water vapor + heat

Evidence of these products being formed may easily be secured. In the first case, if the gas from the burning carbon is bubbled through clear limewater, the water will immediately turn cloudy as insoluble limestone is formed by the carbon dioxide combining with the lime. Nothing except carbon dioxide will so affect the limewater. In the second case, to show that the hydrogen is forming water vapor, all that is necessary is to hold a cold glass over the gas flame, and the water vapor will condense on the glass to form visible drops of water.

Oxygen will combine with many other things, even some metals, rapidly enough for combustion. Magnesium metal in powder or ribbon form burns with a brilliant white light, which makes it very useful as flash powder in making photographic pictures. The white powder formed by the burning process is magnesium oxide.

The rapidity with which oxidation, or burning, takes place depends primarily upon the temperature of the fuel and the amount of oxygen present. In the burning of the fuels just mentioned, the temperatures are quite high. Before burning will take place, it is necessary to raise the temperature of at least a small part of the fuel high enough for oxidation to take place rapidly; *i.e.*, wood, coal, or gas must be "lighted" before they will burn. Some outside heat must be applied to raise the temperature of a small part of the fuel to what is called the "kindling temperature" before rapid oxidation will begin. When the burning is started, the heat of combustion increases the temperature of the remaining fuel to the kindling temperature, and burning continues with the liberation of a large amount of heat.

In some instances oxidation takes place at an exceedingly rapid rate; when it does, the rapid liberation of heat and the enormous expansion produced by the heated gases formed by the oxidation gives us what is called an explosion. The rapid



A grain elevator wrecked by a dust explosion in which a dust cloud inside the concrete storage tanks became ignited. The rapid oxidation produced pressures of over 100 pounds per square inch. (U. S. Department of Agriculture photograph.)

oxidation of gasoline vapor in the cylinder of a gasoline engine is an illustration of this process. In this case the heat, or energy, of the expanding gases is controlled and partly used to do work in driving an automobile or other machines. Similar explosions on a much larger scale occasionally take place in dust clouds formed in granaries or other places of storage. Here the forces are uncontrolled, and they usually produce extensive havoc. Such things as starch, grain dust, wood dust, soap powder, and coal dust when dry and divided finely enough to form a cloud will explode upon being ignited. Ignition may be brought about by an electric spark or other sources of heat which will produce a starting temperature of about 500°C.

In large-scale tests of pressures exerted in dust explosions, it has been found that they may be as high as 200 pounds per square inch. Observations made at industrial plants following explosions have shown that it is impractical to build walls strong enough to withstand the pressures produced. In one case a sixteen-inch reinforced concrete wall was blown out by the

explosion, and in other cases grain elevators of a hundred thousand tons or more have been shattered or moved off their foundations. Within the last twenty years some three hundred persons have been killed and thirty-five million dollars' worth of property destroyed by dust explosions. However, in recent years much progress has been made in discovering how to prevent them; and when these preventive measures are observed, the hazard may be materially reduced.

Some substances will start burning at normal room temperature if they are exposed to the oxygen of the air. For example, when white phosphorus is spread over a piece of paper in finely divided form and exposed to the air, the whole mass will suddenly burst into flame. Under these conditions burning has started at room temperature, as phosphorus has a kindling temperature lower than that of air. This is one illustration of what is referred to as "spontaneous combustion." Sometimes old rags soaked with oil suddenly burst into flame when left lying loosely in a nonventilated closet or room. There is nothing mysterious or magical about such an occurrence; it is the operation of the well-known laws of oxidation. A slow oxidation of the oil takes place, which liberates a small amount of heat that cannot escape from the closed room and still air. Gradually the temperature rises, and after a time the kindling temperature of the oil or rags is reached. Then burning begins spontaneously. Spontaneous combustion occurs for the same reasons occasionally in undisturbed coal mines, sugar refineries, haylofts, and flour mills.

The animal body is a wonderful mechanism for the oxidation of foods. The oxygen of the air combines with the bluish blood in the lungs to form red oxyhemoglobin in the red corpuscles, and the corpuscles are carried in the blood stream to all the minute tissues of the body. There the pure oxygen is released from the blood and goes to the body cells. In the processes of metabolism within the body cells, oxidation of the living substance of the body cells takes place, for the most part converting it into carbon dioxide and water. The heat released by oxidation of the tissue substances maintains the temperature of the animal body, and the amount of heat thus obtained is approximately equal to that secured by burning in air the same amount of food that was used in building up the body tissues.

When a higher percentage of oxygen is present than normal air contains, oxidation takes place much more rapidly. A splint of wood that is burning with only a dull glow in the air bursts into a bright flame when placed in an atmosphere of pure oxygen. Breathing is more easily and rapidly accomplished when one is in an atmosphere of pure oxygen. This is the reason for placing under an oxygen tent a patient with severe lung trouble. It has been reported that mental work is accomplished more accurately and rapidly when one is in an atmosphere of increased oxygen content than when in normal air. Since the nerve cells require more oxygen than any other cells of the body, this is probably true.

Patient in the controlled atmosphere of an oxygen tent. (Courtesy of Linde Air Products.)

In some instances oxidation takes place very slowly. One important illustration is iron rusting. Here the iron combines with the oxygen of the air and with moisture to form an oxide of iron, with the liberation of a small amount of heat. This is the great industrial fault of iron, for millions of dollars worth of iron and steel structures deteriorates annually by the process of oxidation into the red powder, hydrated iron oxide. Man exerts quite extensive and expensive efforts in his attempts to prevent this slow process of oxidation. To protect iron from this chemical reaction it is painted or made into "rustless" or "stainless" steels by mixing it or covering it with other metals to prevent it from coming in contact with the oxygen and moisture of the air.

In the examples noted above, the agent of oxidation is the element oxygen; however, other elements besides oxygen have the ability to act as oxidizing agents. Nature knows no monopoly here. When a burning candle is introduced into a vessel filled with chlorine gas, the candle continues to burn with a smoky flame, and the following reaction takes place:

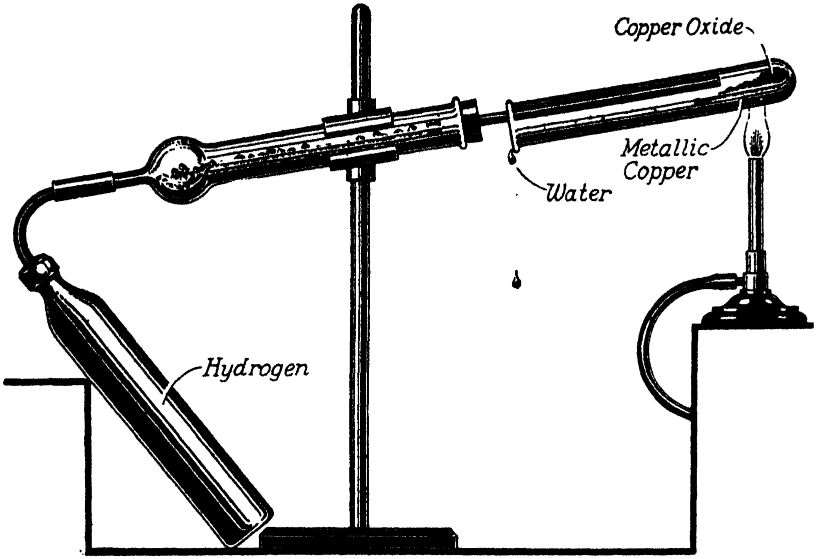
$$\text{Candle} + \text{chlorine} \rightarrow \text{hydrogen chloride} + \underset{\text{(smoke)}}{\text{carbon}} + \text{heat}$$

Likewise, copper burns in chlorine, and a heated iron wire sparkles in this gas much as it does in oxygen. The burning here is visible evidence of the union of these burning substances with chlorine. The products of combustion are chlorides, however, instead of oxides. Iron will also burn with sulphur when the two are mixed in finely divided form and heated to the kindling temperature, as previously noted, producing iron sulphide. Many other substances behave in this same manner to bring about oxidation; *i.e.*, they combine with certain of the elements to form simple compounds of those elements with the reaction releasing a sizable amount of energy. Elements that behave thus are called oxidizing elements, or oxidizing agents. All these reactions are examples of oxidation, as the term is defined in a more general sense.

Reduction, whereby Energy Is Stored

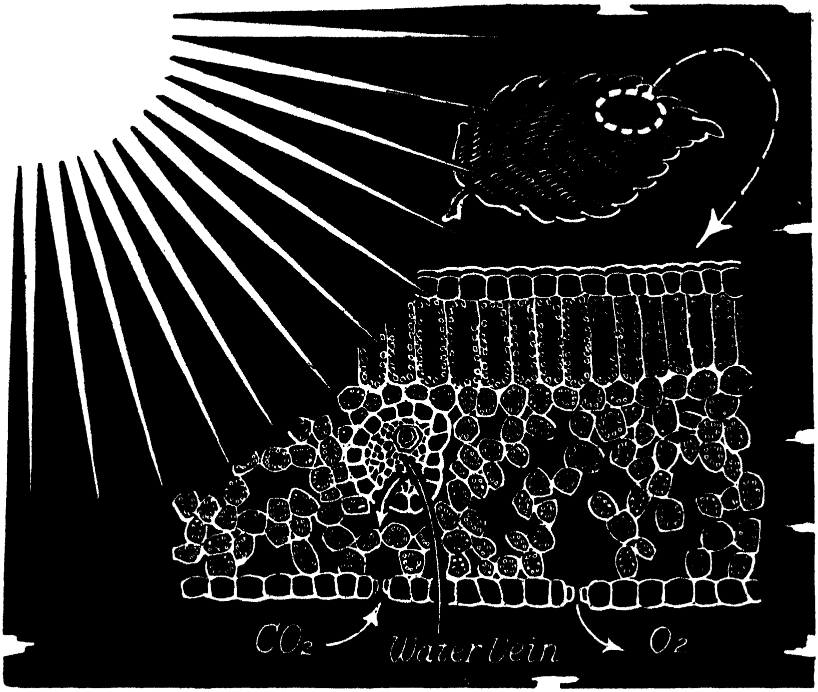
Almost everyone knows that hydrogen combines rapidly and explosively with oxygen when the two free gases are mixed and ignited. The burning of the Zeppelin *Hindenburg* was a tragic example. Hydrogen also combines with oxygen under less spectacular conditions, in which it takes oxygen away from some of its compounds. Here we have one element injecting itself into a chemical compound, uniting with one of the members of the union, and liberating the other one. In this respect it might be compared to the third party of "the eternal triangle" that we have heard so much about in human social relations. For example, when dry hydrogen gas is passed over copper oxide, and heat added, the black oxide turns to red metallic copper, and the hydrogen forms water vapor with the oxygen thus secured. The following reaction occurs:

$$\text{Copper oxide} + \text{hydrogen} + \text{heat} \rightarrow \text{metallic copper} + \text{water vapor}$$



Reduction of copper oxide by hydrogen to metallic copper. It is necessary that heat be applied for the reaction to occur.

Other substances besides hydrogen have the ability to remove oxygen from its compounds. In the metallurgy of iron it is necessary to remove the oxygen from the ore, as most of the iron ores mined are in the form of oxides of the metal. The oxygen is removed on a large scale in the blast furnace by causing carbon (coke) to combine with the oxygen of the ore to form carbon dioxide and leaving the iron free as metallic iron. Before this reduction can be accomplished, it is necessary to add enormous quantities of heat energy to the reacting materials by maintaining a high temperature in the blast furnace. Many other illustrations of this same type of reaction might be cited. The process of combining with oxygen or of removing oxygen from its compounds is known as reduction, and the elements that accomplish it are called reducing agents. It is seen that reduction is just the opposite of oxidation. As illustrated here, reduction is the process of removing oxygen from its compounds when energy is added; whereas oxidation, as described above, is the process of forming oxides or similar compounds with the liberation of energy. It should be noted, however, that reduction in its broader sense has a meaning wider than that given here.



Reduction of carbon dioxide and water by the energy of sunlight to form starch and oxygen in the chlorophyll bodies of green plants.

Just one illustration of reduction operating on a grand scale in nature is the manufacture of foods by living plants, the only ultimate source of food and of many fuels on the earth. This is a synthetic chemical reaction in which carbon dioxide and water are reduced in the chlorophyll bodies of the green leaves of plants to form sugar and other carbohydrates while the extracted oxygen molecules are diffused into the air and thereby support animal breathing. The energy necessary for this reduction reaction is furnished by sunlight. Work is done in the formation of the carbohydrates, and thus power is stored for future use.

When oxidation occurs, heat energy is released; reduction usually requires that heat energy from an outside source be added. To those who would stop to consider, it is evident that oxidation and reduction are extremely important types of chemical reactions. Oxidation is our chief source of power;

reduction is the great method of power storage. Boilers in industrial plants are steamed up to high working pressures by the oxidation of fuels in the furnace, and normal body temperature is maintained in the same manner. Automobiles move over the road and airplanes fly through the air by means of the energy released by the oxidation of gasoline in the cylinder of the gas engine. Most of this energy has been stored in fuels and plant and animal tissues by the chemical processes of reduction, to be released when needed.

A Universal Shorthand System

The expression of chemical reactions as used in the foregoing equations is cumbersome and inexact. As the idea developed in the history of chemical research that reactions took place between individual atoms, a simpler and more exact system of expressing them was devised. This may be thought of as the chemist's shorthand. Standard practice at present is to represent the atom with a letter, which also indicates the chemical element as well as its atom. This letter is known as the symbol for that element, and usually the first letter of the name of the element, expressed as a capital, is used. For example, the symbol for hydrogen is H; for oxygen, O; for carbon, C; for sulphur, S. Where the single-letter representation is found inconvenient because of two names having the same first letter, the first letter of one of the elements is written as a capital followed by another letter of the word written small, such as Cl for chlorine, Ra for radium, and Ni for nickel. In some cases one or more letters of the Latin name of the element are used, as K (kalium) for potassium, Ag (argentum) for silver, and Au (aurum) for gold.

When a chemical compound is to be expressed in this shorthand system, a formula is used. This is secured by combining the symbols of the elements that entered into the reaction; thus, CuO refers to a compound of copper and oxygen, known as cupric oxide. This formula implies that one atom of the metal copper has combined with one atom of oxygen gas to form one molecule of the black powder cupric oxide. The formula for sulphuric acid is H_2SO_4, stating that the molecule of this acid contains atoms of hydrogen gas, sulphur, and oxygen gas

combined in the ratio of 2 to 1 to 4. A formula, then, implies not only what atoms went into the formation of a compound but also the proportion of the different atoms used. A formula gives an exactness to expressing the composition of a compound that is impossible to give by merely writing its name.

A chemical reaction is represented by an equation. Equations may sometimes seem forbidding, yet in reality they are usually, at least in chemistry, a simplified, and thus easier and clearer, manner of expressing exactly chemical reactions. For example, the formation of table salt (sodium chloride) from hydrochloric acid and sodium hydroxide is given by the equation

$$\underset{\text{hydrochloric acid}}{HCl} + \underset{\text{sodium hydroxide}}{NaOH} \rightarrow \underset{\text{sodium chloride}}{NaCl} + \underset{\text{water}}{H_2O}$$

Here a molecule of hydrochloric acid (HCl) combines with a molecule of sodium hydroxide (NaOH) to form a molecule of salt (NaCl) and one of water (H_2O).

Atomic Weights

Before a group of symbols could be put together to represent a compound or be used in writing the equation for a chemical reaction, it was necessary to know how many atoms of the elements were involved. The only way to secure this knowledge was to make an exact analysis of each compound studied. Since it is impossible to see or to count the atoms entering into a reaction, the idea of weight measurements offered a satisfactory means of reducing to some order the chemical chaos of earlier centuries. About the beginning of the nineteenth century the principle of the definite composition of a compound was discovered, and modern highly accurate chemical measurements demonstrate the correctness of the findings. This means that the composition of a pure compound is always precisely and exactly the same. For example, pure water, regardless of the conditions imposed upon it, always contains 11.19 per cent by weight of hydrogen and 88.81 per cent by weight of oxygen. Similar exactnesses have been found in the composition of hundreds of thousands of other compounds.

Let us see how this works out regarding the weights of the atoms making up the compound, using water as an illustration.

It is well known that twice as much hydrogen as oxygen by volume enters into combination to form water, and it has long ago been established that equal volumes of gases contain equal numbers of atoms or molecules. This being true, twice as many atoms of hydrogen as oxygen unite when water is produced. Accordingly, the formula is written H_2O. The weight of the hydrogen atom as compared to that of oxygen will be in the ratio, therefore, of 5.59 (11.19 ÷ 2) to 88.81, or approximately 1 to 16. On this basis the relative weights of the atoms of these two gases were established. The atomic weight of hydrogen was called 1, while that of oxygen was 16.

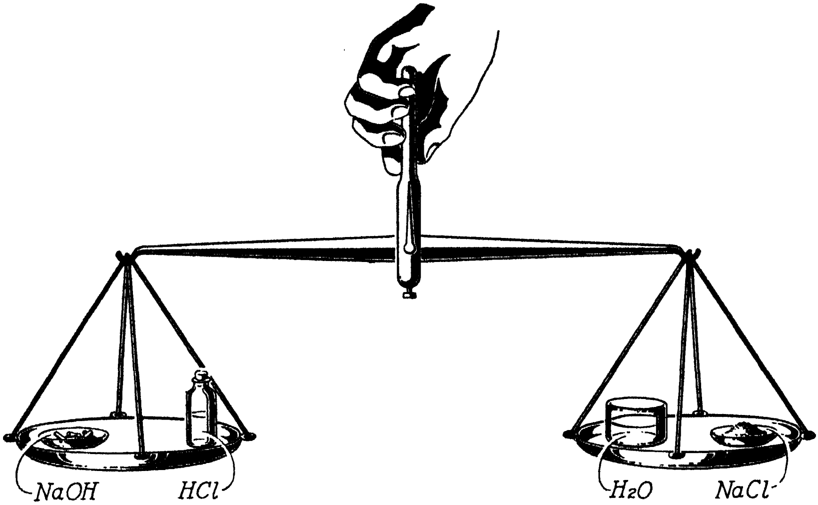
The hydrogen atom is the lightest of the known atoms, and its relative weight of 1 was originally taken as a standard. However, in recent times oxygen has been considered a more satisfactory standard, and it is now used, its arbitrary atomic weight being assigned as 16. On the basis of more refined measurements, it is now known that the exact ratio of the hydrogen atom weight to the 16 for oxygen is slightly over 1, or 1.0078. These figures, then, represent the atomic weights of these two gases. This does not mean that the hydrogen atom weighs 1.0078 grams and oxygen 16 grams or pounds or tons, but it does mean that the weights of their atoms are in this ratio. The numerical values of their actual atomic weights are very small.

When a given number of atoms of oxygen or hydrogen can be made to combine with an equal number of atoms of other elements, the exact relative weights of the other atoms may also be determined by a system of weight comparisons. By such a procedure it is found that the sulphur atom, for example, is about twice as heavy as the oxygen; the copper atom is approximately four times as heavy; and the uranium atom is the heaviest known at present. To be exact, the sulphur atom has an atomic weight of 32.06, and the atomic weight of the copper atom is 63.57. The atomic weights of all the elements have been obtained, and they are published in international tables of atomic weights.

Since a compound is always composed of a definite and constant number of atoms and these atoms have definite and constant weights, the composition of any compound is always

Table of the Elements
Showing Atomic Weights

	Symbol	Atomic number	Atomic weight		Symbol	Atomic number	Atomic weight
Aluminum	Al	13	26.97	Molybdenum	Mo	42	95.95
Antimony	Sb	51	121.76	Neodymium	Nd	60	144.27
Argon	A	18	39.944	Neon	Ne	10	20.183
Arsenic	As	33	74.91	Nickel	Ni	28	58.69
Barium	Ba	56	137.36	Nitrogen	N	7	14.008
Beryllium	Be	4	9.02	Osmium	Os	76	190.2
Bismuth	Bi	83	209.00	Oxygen	O	8	16.0000
Boron	B	5	10.82	Palladium	Pd	46	106.7
Bromine	Br	35	79.916	Phosphorus	P	15	30.98
Cadmium	Cd	48	112.41	Platinum	Pt	78	195.23
Calcium	Ca	20	40.08	Potassium	K	19	39.096
Carbon	C	6	12.010	Praseodymium	Pr	59	140.92
Cerium	Ce	58	140.13	Protastinium	Pa	91	231
Cesium	Cs	55	132.91	Radium	Ra	88	226.05
Chlorine	Cl	17	35.457	Radon	Rn	86	222
Chromium	Cr	24	52.01	Rhenium	Re	75	186.31
Cobalt	Co	27	58.94	Rhodium	Rh	45	102.91
Columbium	Cb	41	92.91	Rubidium	Rb	37	85.48
Copper	Cu	29	63.57	Ruthenium	Ru	44	101.7
Dysprosium	Dy	66	162.46	Samarium	Sm	62	150.43
Erbium	Er	68	167.2	Scandium	Sc	21	45.10
Europium	Eu	63	152.0	Selenium	Se	34	78.96
Fluorine	F	9	19.00	Silicon	Si	14	28.06
Gadolinium	Gd	64	156.9	Silver	Ag	47	107.880
Gallium	Ga	31	69.72	Sodium	Na	11	22.997
Germanium	Ge	32	72.60	Strontium	Sr	38	87.63
Gold	Au	79	197.2	Sulphur	S	16	32.06
Hafnium	Hf	72	178.6	Tantalum	Ta	73	180.88
Helium	He	2	4.003	Tellurium	Te	52	127.61
Holmium	Ho	67	163.5	Terbium	Tb	65	159.2
Hydrogen	H	1	1.0080	Thallium	Tl	81	204.39
Indium	In	49	114.76	Thorium	Th	90	232.12
Iodine	I	53	126.92	Thulium	Tm	69	169.4
Iridium	Ir	77	193.1	Tin	Sn	50	118.70
Iron	Fe	26	55.85	Titanium	Ti	22	47.90
Krypton	Kr	36	83.7	Tungsten	W	74	183.92
Lanthanum	La	57	138.92	Uranium	U	92	238.07
Lead	Pb	82	207.21	Vanadium	V	23	50.95
Lithium	Li	3	6.940	Xenon	Xe	54	131.3
Lutecium	Lu	71	174.99	Ytterbium	Yb	70	173.04
Magnesium	Mg	12	24.32	Yttrium	Y	39	88.92
Manganese	Mn	25	54.93	Zinc	Zn	30	65.38
Mercury	Hg	80	200.61	Zirconium	Zr	40	91.22



Weight values are maintained in chemical reactions.

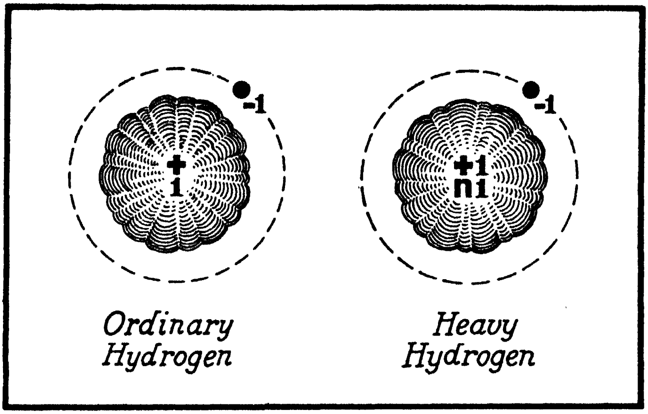
exactly the same. This permits, then, of an accuracy of weight values being expressed in a chemical equation. In the neutralization of hydrochloric acid with sodium hydroxide, as noted above, one atom of chlorine of weight 35 combines with one atom of sodium of weight 23 to form a molecule of salt of weight 58. One atom of hydrogen weighing 1 combines with one radical molecule (OH) weighing 17 to form a molecule of water weighing 18. This equation may now be more precisely written:

$$\begin{array}{ccccccccc} HCl & + & NaOH & \rightarrow & NaCl & + & H_2O \\ (1+35) & + & (23+16+1) & = & (23+35) & + & (2+16) \\ 36 & + & 40 & = & 58 & + & 18 \\ & & 76 & = & 76 & & \end{array}$$

Furthermore, the total weight of all substances formed is exactly the same as the total weight of all substances that entered into the chemical reaction. Thus, the weight values give an exactness and definiteness to chemical reactions which are important in understanding chemical changes.

Double Personalities

Most of the chemical elements are made up of atoms with a sort of Dr. Jekyll-Mr. Hyde complex. What is implied by this is that many elements have scatterings of atoms that are heavier or lighter than the normal ones, but in all other respects these other members are exactly like the normal atoms. Such

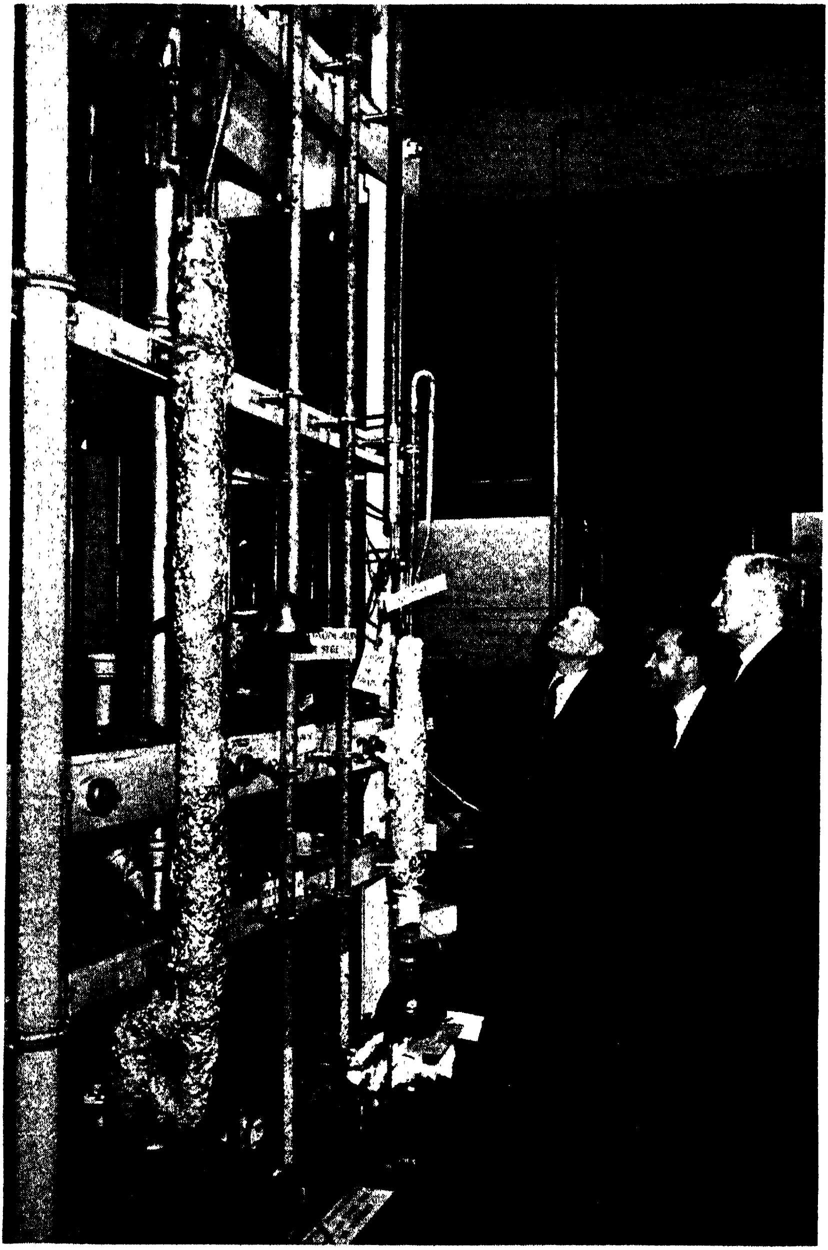


Graphic representation of ordinary and heavy hydrogen atoms.

heavy or light atoms of a given element are called isotopes. We might say that the isotope of an element is really a different element but that it has the same chemical name, the same number of outer electrons, and the same chemical properties. It has a different atomic weight, however, because of the existence of more or fewer neutrons in the nucleus. This makes the nucleus of the isotope have a different weight, although the atom has the same number of electrical charges and, therefore, the same set of chemical properties.

A highly publicized example of this is heavy hydrogen, a hydrogen atom with a proton and a neutron in the nucleus instead of just a proton, as in the case of ordinary hydrogen. This hydrogen isotope has an atomic weight of approximately two as compared to roughly one for ordinary hydrogen. However, its nuclear charge and its outer electron are the same as they are in regular hydrogen, which gives heavy hydrogen the same chemical properties as ordinary hydrogen. The hydrogen isotope, discovered by Dr. Harold C. Urey of Columbia University (and for which he received the Nobel Prize in 1933), is present in rare traces in ordinary water. Heavy water is the popular expression used in referring to molecules of water containing heavy hydrogen atoms.

Heavy water is now prepared commercially, so great is the present demand for it for scientific research. It is prepared by a continued process of electrolysis of ordinary water, and by such



Dr. Harold C. Urey of Columbia University (left) demonstrates his 20-foot high glass apparatus for the isolation of heavy carbon to Prof. Enrico Fermi (center) and Dean George B. Pegram. (Courtesy of the Science Observer.)

a process it is now possible to prepare 99 per cent pure heavy water. Heavy water is used extensively to provide particles for bombardment of other atoms to produce transmutation. It furnishes particles, *viz.*, neutrons and protons, par excellence for this purpose when they are actuated with a million volts of electricity or more in a cyclotron. The speed of the emitted particles is proportional to the voltage applied, and it is possible with this method to send 2×10^{15} particles per second into a target at speeds up to 10,000 miles per second. When such large numbers of atomic particles are sent flying into the nuclei of atoms at these enormous speeds, a transmutation of the atoms bombarded is accomplished on a relatively large scale. Heavy water is also used extensively in many bacteriological and biological investigations.

Chlorine is another instance in the case of isotopes. This element really consists of two different atoms: one with a nucleus of 17 protons and 18 neutrons, the other with a nucleus of 17 protons and 20 neutrons. The two atoms have, therefore, atomic weights of 35 and 37, respectively. Ordinary chlorine consists of 77 per cent of the isotope of atomic mass 35 and 23 per cent of the isotope of atomic mass 37. The mean mass of ordinary chlorine is 35.46, as given in the Table of the Elements. Since each of these isotopes contains the same number of outer electrons, namely, 17, they have the same chemical properties. There is no chemical way by which the two isotopes may be either identified or separated from each other. Accordingly, ordinary chlorine gas has one set of chemical properties even though it is made up of atoms of two different atomic weights. These atoms are identified and their relative percentages measured by processes that involve weight characteristics rather than by chemical processes.

Many other elements contain isotopes; in fact isotopes have been discovered in nearly all the ninety-two elements. Some elements have six or eight different isotopes. The metal platinum is an illustration, since it is composed of isotopes with atomic weights of 192, 194, 195, 196, and 198 distributed throughout the element in such proportions that the combined atomic weight is 195.23. The existence of isotopes in the elements explains why their atomic weights have fractions rather than being

whole numbers. Copper, for example, has an atomic weight of 63.57 and consists of two isotopes, one with an atomic weight of 63 and the other with an atomic weight of 65. The fractional part of its combined atomic weight is a measure of the percentage of the heavier isotope present.

Reaction Velocity

Anyone who has ever lived for a time in the "open spaces" away from city lights appreciates the value of flashlights at night. The energy supplied by the chemical reactions within the battery is easily used to light the way in the darkness. Admiral Byrd, however, found that he could not use flashlights in the South Polar regions because the extreme cold stopped the necessary chemical action of the dry batteries. Those who have ever had the pleasure of developing photographic films know that the temperature of the developer affects the rate at which development takes place. One kind of developer, for example, requires thirty minutes at 55°F. and fifteen minutes at 70°F.; thus, raising the temperature 15° makes the reaction work twice as fast. With a great many compounds it has been roughly approximated that increasing the temperature 10°F. about doubles the speed of reaction.

This is evidence that in many instances heat plays an important part in either hastening a chemical reaction or bringing it about. In general heat is important under conditions where the reacting substances are all in the gaseous state or where they are all in the liquid state or all are in solution. In order for atoms and molecules to combine to form compounds they must collide with extreme atomic and molecular forces. The atoms must approach each other so closely that their attractive forces are great enough to hold them together in compounds. By heating two reacting gases, the atoms and molecules are speeded up; their "force of action" becomes greater; and their union in compounds takes place or is greatly accelerated. It probably is generally known that hydrogen and oxygen gases may be mixed and remain together indefinitely at atmospheric temperature without reacting; when, however, their temperature is increased by an electric spark or a flame, they combine with explosive violence.

Chemical landscape produced by formation of tiny fernlike crystalline masses from cholesterol dissolved in alcohol. The insert shows the crystalline formation in the beaker, and the artistic landscape effect is secured from an enlarged negative. (Photograph by Clayton F. Smith.)

Substances in solution or in liquid form present an ideal situation for chemical reactions to occur. The volume is very compact as compared to the gaseous state, and yet the molecules are freely mobile. They may come in contact with each other readily; and if they have any disposition to combine, they are most likely to do so. In such cases an increased temperature greatly accelerates the reaction velocity by speeding up the motion of the molecules so that they approach each other with greater forces. Conversely, lowering the temperature slows down the molecular motion, and the rate of reaction is decreased or entirely stopped.

On the other hand, when a gaseous substance is dissolved in a solution, any chemical reaction that might occur between the gas and the solution is likely to be retarded by increasing the temperature. Let us see why this should be so. When heat is

applied, the speed of the gas molecules is greatly increased, and as a result some of the gas molecules escape from the solution, just as water molecules escape as a gas (water vapor) when water is boiled. When some of the gas molecules are driven out of the solution, fewer remain to form a compound, and the reaction is thereby slowed up. However, if an increased pressure is applied rather than heat, the gas molecules are forced to stay in the liquid; their molecular speed is increased, involving more chances of effective collision; and the reaction velocity is aided thereby.

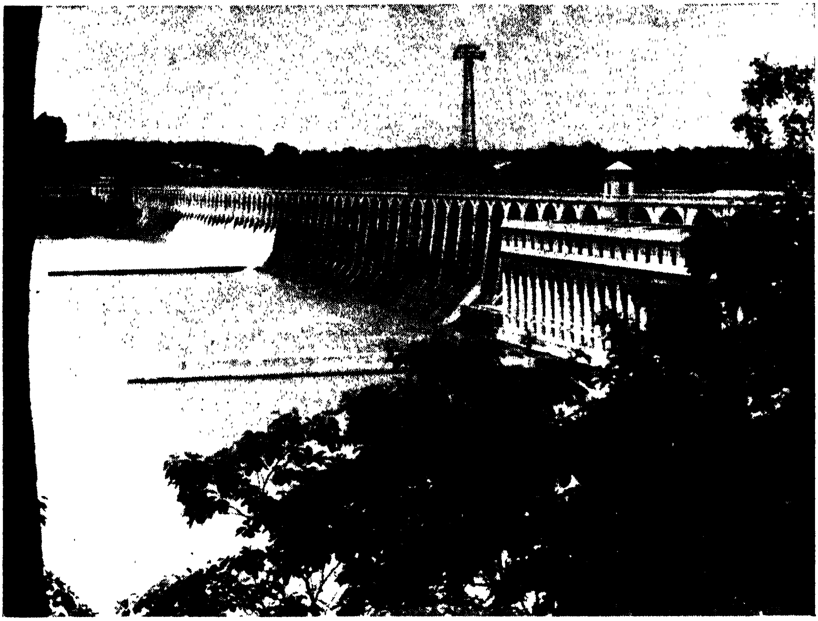
The velocity of a great many reactions may be increased or decreased by what is known as a catalyst. By a catalyst is meant a substance that influences the speed of a chemical reaction without itself permanently entering into the reaction. Two examples may serve to give some understanding of how a catalyst influences chemical reactions and the importance of catalysts in modern industry.

Consider first the laboratory method of preparing oxygen gas. Potassium chlorate ($KClO_3$) is placed in the proper apparatus and heated; but when it is alone in the apparatus, even if it is heated until it melts, oxygen is evolved only slowly. However, when the black powder manganese dioxide (MnO_2) is added to the potassium chlorate and the mixture heated to much lower degree, oxygen is liberated freely. A chemical analysis of the residue shows that the manganese dioxide is unchanged and that the oxygen came from the potassium chlorate. It may be concluded, therefore, from this reaction that the catalyst does not permanently enter into the reactions and also that the presence of the catalyst demands less energy for the reactions to proceed. Other tests conducted in a highly accurate manner have shown additional characteristics of catalysts in chemical reactions. One of them is that the more finely divided the catalyst is and the more intimately it is mixed the better are the results obtained. Another is that the quantity of the catalyst may be exceedingly small as compared to the quantity of the reacting substances and that the catalyst may be used over and over again provided it does not become contaminated with a foreign substance.

The second example that we wish to mention applies to the importance of catalysts in modern industry. One important industrial situation that developed during the first World War in which a catalyst played a leading role is the production of nitrogen compounds from atmospheric nitrogen. One of the most valuable groups of compounds is the nitrogen compounds. They are important parts of foods, explosives, fertilizers, motion-picture films, not to mention a host of others. They are found plentifully distributed in nature only in a few places, chiefly in Germany and Chile, with over half the world supply in Chile. Until recent times the nations of the earth were dependent primarily upon Chile for their supply of commercial nitrates. However, the first World War disrupted this commercial setup; no longer were the natural nitrates available to all nations.

The most plentiful substance of the air is nitrogen, but under ordinary conditions it does not form compounds. Again the old problem was "Water, water everywhere, nor any drop to drink." At about the turn of the present century it was discovered in Germany that nitrogen may be made to combine directly with hydrogen to form ammonia (NH_3). The ammonia secured in this manner may be used to make other nitrogen compounds. Now some of the greatest industrial plants that man has constructed operate for the purpose of producing ammonia from atmospheric nitrogen.

One method of doing this is to bring together nitrogen and hydrogen under conditions favorable to their combination and in the presence of that touchstone of modern chemistry, a catalyst. The process was invented by Prof. Fritz Haber of Germany in 1913, and the Haber process, or a variation of it, is now used extensively in many countries. The nitrogen and hydrogen gases are brought together under pressures of 200 atmospheres (3,000 pounds per square inch) or greater, and the temperature raised from 300 to 500°C. The pressures cause a realignment of the atoms of the two gases, and the action of the catalyst facilitates their union to form ammonia. The catalyst employed is usually finely divided metallic iron. Without it, the gases will not react in any profitable quantities; with it, as much as



The Wilson Dam in Alabama for the production of power to synthesize ammonia from the nitrogen of the air. (Wide World photograph.)

85 per cent production of ammonia is secured in some of the processes. As a result of this discovery and its application in various forms, the world production of commercial nitrogen compounds increased from less than one million tons in 1913 to over three million tons in 1938.

The speed of chemical reactions may be controlled in still another manner. This is by regulating the concentrations of the reacting substances. Let us keep in mind that reactions take place between individual atoms or molecules and that they cannot act upon each other at a distance, and it will be evident that any condition that increases the frequency of molecular collision will promote the speed of the reaction. The more molecules there are in a given space the faster the reaction will proceed. Hence, an increase in concentration will increase the reaction velocity, and a decrease in concentration will produce a reduction in this velocity. For example, substances burn more brilliantly in pure oxygen than in air. We know that only one-fifth of the air is oxygen; consequently, pure oxygen at the same

pressure has five times the concentration of air. This additional concentration of oxygen molecules is what produces the more rapid burning.

Thus, by various methods, chemical reactions may be started, speeded up, slowed down, or stopped. Such control being known to man has enabled him to govern or develop many important industrial processes as well as to bring about many medical practices that have been of unsurpassed economic and social value.

REFERENCES FOR MORE EXTENDED READING

GRADY, RAY I., *et al.*: "The Chemist at Work," *Journal of Chemical Education*, Easton, Pa., 1940.

Each of fifty-two successful chemists working in many branches of the science has described his particular work with candor and interesting detail.

FINLAY, ALEXANDER: "A Hundred Years of Chemistry," The Macmillan Company, New York, 1937.

An abbreviated but informative history of chemistry since the first quarter of the nineteenth century, including biographical sketches of the great chemists of the period.

FURNAS, C. C.: "The Next Hundred Years," The Williams & Wilkins Company, Baltimore, 1936.

Part II of this thought-provoking and interestingly written book presents some of the immediate problems relating to chemistry that face mankind at present. The four years since its publication have witnessed achievements along certain lines far beyond the author's analysis.

HATCHER, W. H.: "An Introduction to Chemical Science," John Wiley & Sons, Inc., New York, 1940.

The best organized and most clearly and concisely written text for a course in chemistry for students in branches of learning other than the sciences that has come to this reviewer's attention. Intelligent use of the index will enable the reader to locate a discussion of most of the fundamentals of chemical science.

HOLMES, HARRY N.: "General Chemistry," The Macmillan Company, New York, 1936, Chaps. I–V, VII, XXIV, XXV.

This standard college textbook in general chemistry combines accuracy with clear statement. Most of the topics discussed are illustrated with fundamental and typical chemical reactions. The general reader will find in the above-noted chapters introductory chemical concepts and some discussion of atomic structure and nuclear chemistry.

RICHARDSON, L. B., and A. J. SCARLETT: "General College Chemistry," Henry Holt & Company, Inc., New York, 1940.

The first ten chapters of this excellent college text present in a type of organization and style of writing readily understandable to the beginning student an introduction to many of the basic chemical concepts.

ARTHUR, PAUL: "Lecture Demonstrations in General Chemistry," McGraw-Hill Book Company, Inc., New York, 1940.

Those wishing complete instructions for presenting lecture demonstrations in the classroom or auditorium will find the many hundreds of chemical experiments here described a valuable help.

New Edition, published by The American Chemical Society, Easton, Pa.

A weekly journal of current news of chemistry and chemical engineering that is probably the best digest of chemical activities published. Popularly written articles, well illustrated.

Journal of Chemical Education, published by The American Chemical Society, Easton, Pa.

Articles relating to the history and use of chemistry as well as to chemical research, usually written in nontechnical language.

Science Service.

6: COMBINING AND SEPARATING

Or the Nature of Chemical Changes

JUST how atoms combine to form compounds has been a baffling question. At one time the belief was that they were held together by means of small hooks, or barbs. At other times such changes were accounted for by supposing that the atoms had different shapes which allowed them to fit together in an orderly manner, perhaps in a way not unlike our modern jigsaw puzzles. Both these assumptions are now known to be untrue. The chemists of the earth have continued an unceasing investigation into the nature of chemical processes during the last century, and every known source of possible information has been explored. The early speculations regarding the nature of chemical changes have given way to ideas based upon more

reliable scientific data. Although our present ideas may as yet be incomplete, they do conform to a large body of experimentally determined facts.

It was discovered about a century ago that the elements were not entirely dissimilar in their chemical properties, and still later it was found that they could be classified into a small number of groups according to the compounds that they formed and according to their chemical behavior. The explanations of chemical changes, then, were explanations that could be applied to elements in groups. It was also discovered that the compounds of some elements when dissolved in water formed solutions that had the property of conducting an electric current. Such solutions are known as electrolytes, and apparently the dissolved compounds become broken down into electrically charged particles known as ions. Furthermore, it became known that the compounds producing nonelectrolytic solutions never form ions. Here, then, was another clue that offered possibilities of acquiring an understanding of the nature of chemical changes.

Let us, then, review some of the evidence and ideas regarding the nature of chemical changes and glean from them the best possible understanding of how these changes take place and the significance of some of them in our daily lives. The presentation given here must be both selective and brief. It is not our purpose to consider the wealth of detail, which is both interesting and instructive, but rather to point out, as well as possible, the most significant factors and relationships.

Chemical Families

Man may occasionally think that family relationships are peculiar to mankind, yet among the chemical elements are groups with similar characteristics and common properties. These groups constitute the chemical families. It has been of great value to man in the study of chemistry that the elements are not entirely unlike. The way to a fuller understanding of the nature of chemical change has been made simpler; research has thereby been greatly stimulated. The existence of families of elements with related properties points to the operation of a fundamental law regarding the chemical properties of the elements, and this law has now been clearly established.

Since some of the elements have similar properties, it is possible to place them in groups on the basis of these similarities. This grouping has a close relationship to the atomic weights of the elements and a still closer relationship to the atomic numbers.

Dmitri Mendeléeff, from the Lyman C. Newell History of Chemistry Collection.

The relationship of chemical properties of certain elements to their atomic weights was first observed in 1829 by the chemist Döbereiner, when he found that several groups of three elements each could be so arranged that the atomic weight of one was the average of the other two and that in all such cases the three elements strikingly resembled each other in their properties. Some thirty-five years later an English chemist, John Newlands, announced a more far-reaching discovery of this type of relationship. He had arranged in order of their increasing atomic weights all the elements known at that time and found that "the eighth element, starting from a given one, is a kind of repetition (in properties) of the first, like the eighth note of an octave in music." He read his paper before the Chemical Society of London in 1864, but it was so severely ridiculed by his fellow chemists that he dropped the idea.

Within five years after this meeting the great Russian chemist Dmitri Mendeléeff, entirely unacquainted with Newland's work, outlined the periodic relations much as we know them today. His simple statement of the law was, "The whole of the properties of the elements, both chemical and physical, vary in a periodic fashion with their atomic weights." The substance of this statement is that when the elements are placed in a table in the order of their atomic weights, they vary in their properties periodically as the atomic weight increases. In this periodic

variation, the properties repeat themselves to a certain extent. In his statement of the law, Mendeléeff did not limit the cycle of variation to eight; since the later discovery of the inert gas elements, however, many of the cycles do include eight elements. The relationship represents one of the most fundamental in nature, and its discovery is a monument to Mendeléeff's work and basic thinking as well as a rebuke to the ridicule of less capable minds.

So exact were Mendeléeff's observations that it was possible for him to predict the discovery of a number of unknown elements and to describe their properties. The elements were subsequently discovered, and their properties as predicted by Mendeléeff proved remarkably accurate. However, when still other elements were discovered, certain discrepancies in this classification developed that could not be explained in any way by the known facts at hand. For example, when the element iodine was discovered and placed in the table according to its atomic weight, it did not fall into its proper chemical family but rather into a family that had properties entirely dissimilar to it. Also, there were a few others. Apparently, some other characteristic of the atoms closely associated with atomic weights was the key to this variation. It remained for a young British scientist, Henry Moseley, to discover what the fundamental relationship is.

What Moseley did was to determine accurately the number of extranuclear electrons in the atoms of the different elements. This constitutes what is now known as the atomic number of the elements. He accomplished this measurement by a study of the X-ray spectra of the atoms. Moseley found that in passing from one element to another the X-ray lines shifted and that this shift was a measure of the number of outer electrons. Starting with hydrogen, the elements were numbered according to this shift. Hydrogen was given the number one, helium was number two, and so on for each of the other elements. The number so obtained is the atomic number. There is a fundamental difference between the meaning of atomic weight and atomic number. The atomic weight of an element refers to the mass of each of its atoms, and it is determined primarily by the structure and composition of the nucleus within the

atom; the atomic number is a measure of the electrical particles outside the nucleus, and it is the same as the number of extranuclear electrons.

Henry Moseley, from the Lyman C. Newell History of Chemistry Collection.

The elements were then arranged in the Periodic Table according to their atomic numbers. It was found that the new arrangement placed the elements in the same order in the Periodic Table as they had been placed according to their atomic weights, except for the elements that had been placed in the wrong family by their atomic weights. These elements were placed in their proper groups in the Periodic Table when arranged according to their atomic numbers. The shift from atomic weights to atomic numbers as a basis of determining the chemical properties of the elements is highly significant, since it recognizes the atom as consisting primarily of electric charges. Atomic weights involve mainly the mass of the nucleus and fail to take into account the extranuclear electric charges as important parts of the atoms that influence their chemical behavior. The number and arrangement of these extranuclear electrons are now known to constitute the important factor in determining how the atom will unite with other atoms.

Thus, a young physicist who met death at the age of twenty-eight from an enemy bullet while fighting in the British army at Gallipoli in 1915 had just previous to the outbreak of the first World War determined exactly this fundamental and

significant key to the properties of the chemical elements. It will ever be a monument to his name.

Periodic Variation

The chemical families are determined by this periodic variation of their properties as the atomic number increases. Omitting hydrogen, the first eight elements according to their atomic numbers are helium (2), lithium (3), beryllium (4), boron (5), carbon (6), nitrogen (7), oxygen (8), and fluorine (9), all of which differ markedly from one another. Now let these eight elements along with their atomic numbers be placed in a horizontal row in a table as follows:

He	Li	Be	B	C	N	O	F
2	3	4	5	6	7	8	9

The tenth element, neon, is very similar to helium, and the eleventh element, sodium, is very similar to lithium. Since neon is similar to helium, it may be placed beneath helium in the table, and a second horizontal row begun. This second row includes neon (10), sodium (11), magnesium (12), aluminum (13), silicon (14), phosphorus (15), sulphur (16), and chlorine (17).

Ne	Na	Mg	Al	Si	P	S	Cl
10	11	12	13	14	15	16	17

The eighteenth element, argon, resembles helium and neon; therefore, a third row is started with argon. This same characteristic of the elements is noticed for all the other elements, with some variation, and the table may be continued.

When the elements are arranged in horizontal rows of eight, with some exceptions, and these rows placed one under the other, the vertical columns are made up of elements of more or less pronounced resemblance. Not only are helium, neon, argon, krypton, and xenon very similar in properties, but also other groups are very much alike. Examples of these other groups are lithium, sodium, potassium, and rubidium which make up the sodium family; and beryllium, magnesium, and calcium which compose another family. This variation is called the periodic law, and the grouping of the elements according to this variation forms the Periodic Table. The vertical groups in the table are what make up the families of the elements.

THE PERIODIC TABLE

Group	0	I	II	III	IV	V	VI	VII	Transition triads
Period		H 1							
I	He 2	Li 3	Be 4	B 5	C 6	N 7	O 8	F 9	
II	Ne 10	Na 11	Mg 12	Al 13	Si 14	P 15	S 16	Cl 17	
III	A 18	K 19 Cu 29	Ca 20 Zn 30	Sc 21 Ga 31	Ti 22 Ge 32	V 23 As 33	Cr 24 Se 34	Mn 25 Br 35	Fe Co Ni 26 27 28
IV	Kr 36	Rb 37 Ag 47	Sr 38 Cd 48	Y 39 In 49	Zr 40 Sn 50	Cb 41 Sb 51	Mo 42 Te 52	Ma 43 I 53	Ru Rh Pd 44 45 46
V	Xe 54	Cs 55 Au 79	Ba 56 Hg 80	Rare Earths 57–71 Tl 81	Hf 72 Pb 82	Ta 73 Bi 83	W 74 Po 84	Re 75 85	Os Ir Pt 76 77 78
VI	Rn 86	87	Ra 88	Ac 89	Th 90	Pa 91	U 92		

The first family, composed of the elements at the left side of the Periodic Table, is the helium family. This group is labeled 0 in the Periodic Table. All these elements were unknown in Mendeléeff's times and have later been discovered and added to the table. They are all gases and are sometimes called the "noble gases," since they are extremely inactive and will not "associate" chemically with any of the elements. This means that they never form compounds. Even the atoms of any one element remain aloof from each other and never unite in pairs, as the atoms of most other gases do, to form molecules.

The next group, called Group I in the Periodic Table, includes the sodium family. It is made up of elements quite the opposite of the inert gases. They are all metals, some of which are very active and form compounds readily. One group of these compounds is known as alkalies, and the elements forming them are called the alkali metals. Another distinctive group is the chlorine family which forms a part of Group VII. This family is composed of the elements fluorine, chlorine, bromine, and iodine. These are all active elements which combine easily with other elements to form compounds, in particular a group known as salts.

Family Traits

Within a chemical family there are marked resemblances as well as some variations in both physical and chemical properties of the constituent elements. The resemblances are characteristic of the entire family while the variations usually occur in the same ratio as the increase of the atomic numbers of the elements. These properties may be illustrated with the chlorine family. When the elements are considered in the order of their increasing atomic numbers, fluorine is a lightweight, colorless gas; chlorine is a heavier and yellowish gas; bromine is a liquid which easily forms a heavy, brown vapor; iodine is a solid which sublimes to a dense, purple vapor. With respect to their chemical properties all of them are active elements which enter into the same types of reaction. Fluorine combines with hydrogen with explosive violence; it decomposes water rapidly; and its acid dissolves glass. Hydrofluoric acid is one of the few substances

active enough to react with glass. Chlorine combines readily with hydrogen but with less violence; it reacts with water; and hydrochloric is one of the strongest acids. The chemical conduct of bromine is similar to that of chlorine except that it is less active. Iodine is still less active but it forms a strong acid. When dissolved in alcohol, iodine forms a strong antiseptic, known as "tincture of iodine," for the treatment of wounds.

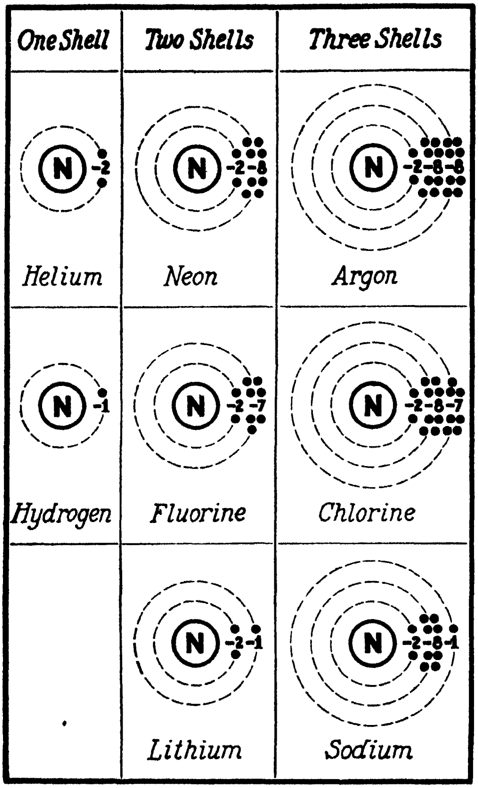
Secret of the Elements' Behavior

The fact that elements group themselves into families and that elements of some families are chemically active while those of other families are inactive is no accident of nature; it is based upon the fundamental way in which chemical changes occur. Just how is it, then, that chemical reactions take place?

The chemical properties of elements and chemical changes are nicely accounted for by the arrangement of the extranuclear electrons, or negative charges, of their atoms. From spectroscopic evidence the electrons surrounding the nucleus appear to be arranged in different "shells" rather than to exist individually or to be scattered in a haphazard manner. The shells are closely related to the different energy levels of electron within the atom, as mentioned in a previous chapter. The innermost shell nearest the nucleus is capable of containing one or two electrons but no more than two. The next outward shell can contain from one to eight electrons but not more than eight. The third has a tendency to be complete with eight electrons; but it may, under certain conditions, contain more than eight, the maximum number being 18. The fourth shell may contain 18, and the fifth 32, electrons; but the arrangements in the fourth and fifth shells are too complex to be considered here. The shells are always filled from the nucleus outward, depending upon the number of outer electrons in the atom.

When the outermost shell containing any electrons is full or complete, the element is inactive chemically. This would be the condition for elements with atoms having 2, 10, 18, 36, and 54 outer electrons. These figures may be verified, if desired, by a little arithmetic on the part of the reader, considering the conditions mentioned in the paragraph above. The elements having these numbers of electrons are helium, neon, argon,

krypton, and xenon, respectively, elements making up the family of noble, or inert, gases. However, when the outermost shell is incomplete, the element is chemically active, and the degree of activity depends somewhat upon the number of electrons missing in the outer shell. The elements that have one electron missing in the outer shell are those having atoms consisting of 1, 9, 17, and 35 outer electrons. These are the very active gases hydrogen, fluorine, chlorine, and bromine. The elements that have only one electron in the outer shell of their atoms have 3, 11, 19, and 37 outer electrons. These elements are the metals lithium, sodium, potassium, and rubidium. They, also, are all very active.



Graphic representation of complete (top row) and incomplete "shells" of atoms.

When elements react chemically, they do so by a transfer between one another of some of the electrons of the outer shells of their atoms when those outer shells are incomplete. There are two distinct and different ways in which this transfer of electrons is made between elements in various types of chemical reactions. In one case there is a "borrowing" and "lending" of electrons by the atoms taking part in the reaction. In the other there is a "sharing" of electrons between the atoms reacting with each other. When the outer shell is incomplete, the atom may borrow electrons from other atoms with incomplete shells or lend electrons to other atoms; or, in the second

case, there may be a mutual sharing of electrons between atoms with incomplete outer shells. This transfer of electrons between atoms with incomplete shells is the bond that holds elements together in compounds. Atoms with complete shells, *i.e.*, the noble gases, do not effect any transfer of electrons; hence they are chemically inactive and do not form compounds.

Valence

It may be surmised from what was briefly stated in the foregoing paragraphs that the extranuclear electrons of significance in chemical actions are those in the outer shell. The various chemical properties of the elements are associated with these outer-shell electrons. They determine the valence of the atoms, which, in turn, is the clue to the exact way the atoms enter into chemical combination. The electrons in the outer shell are usually referred to as valence electrons. In some respects the discovery of valence of the elements was the most important chemical concept of the nineteenth century. Let us note briefly what valence is and how it is related to the electrons in the outer shell of the atoms.

As a great variety of chemical reactions were analyzed during the past century, it was found that in not a single compound did the hydrogen atom combine with more than one atom of other elements. Therefore, it was reasoned that whatever the bond was that held elements together in compounds, hydrogen had only one such bond. Likewise, it was found that chlorine had only one bond in its chemical reactions. A graphic way to represent these two elements is to indicate the one bond with a line, such as H— and Cl—. When these two elements unite, the single bond from each unites, and the compound H—Cl is formed. The oxygen atom was found to combine with two atoms of hydrogen, and accordingly it must possess two of these combining bonds and may be represented as —O—. When this element is combined with hydrogen, each of whose atoms contains a single bond, two hydrogen atoms must be used to complete the oxygen bonds, and the compound may be represented as H—O—H. Many other elements have been found to manifest a double bond when they enter into chemical combinations.

The nitrogen atom combines with three atoms of hydrogen, or their equivalent, when it unites in chemical reactions. A common compound of nitrogen and hydrogen is ammonia (NH_3). Nitrogen must possess, therefore, three of these combining bonds and may be represented as $\begin{array}{c} | \\ -N- \end{array}$, with its hydrogen compound as $\begin{array}{c} H \\ | \\ H-N-H \end{array}$. Carbon, on the other hand, combines with four hydrogens, and a familiar compound of such a union is methane (illuminating) gas, with the formula CH_4. The four bonds for carbon may be represented as $\begin{array}{c} | \\ -C- \\ | \end{array}$, and its hydrogen compound as $\begin{array}{c} H \\ | \\ H-C-H \\ | \\ H \end{array}$. Other elements showed evidences of three or four bonds when they were combined. Still others manifested these combining bonds up to as high as eight.

Thus arose the valence system to account for the exact proportion in which atoms combine with each other when they form compounds. The valence of an atom, then, is the property that determines how many atoms of any other element it can hold in combination or can displace in a reaction. The first four valences may be expressed in summary as follows:

HCl hydrogen chloride	H_2O water	NH_3 ammonia	CH_4 methane
H - Cl	H - O - H	$\begin{array}{c} H \\ \vert \\ H-N-H \end{array}$	$\begin{array}{c} H \\ \vert \\ H-C-H \\ \vert \\ H \end{array}$

Hydrogen and chlorine have valences of 1; oxygen, a valence of 2; nitrogen, 3; and carbon, 4. These and other elements show their valence in whatever chemical union they enter, not being limited to the combination with hydrogen. Carbon dioxide has the well-known formula CO_2 and represented graphically to show the valence of the elements is

$$=C= \quad + \quad O= \quad + \quad =O \quad \text{gives} \quad O=C=O$$

Valence bonds give us an understanding of another well-known property of some elements, particularly those which are gases. This is that even in a pure element in the gaseous state there is a pairing of the atoms to form molecules of the gas. At ordinary temperatures hydrogen does not exist in pure form as individual hydrogen atoms but rather as two hydrogen atoms joined to form hydrogen molecules. The same is true of oxygen, chlorine, nitrogen, and all other pure gaseous elements except the noble gases. This probably means that no atoms can exist in the gaseous state with free or unattached valence bonds. The hydrogen atom does not remain with a free valence bond as H—, nor oxygen with two free bonds as —O—; rather two of the atoms of each element seem to unite with each other so as to form connecting bonds. The gaseous elements are represented, therefore, as

$$\mathrm{H-H} \text{ or } \mathrm{H_2}, \qquad \mathrm{Cl-Cl} \text{ or } \mathrm{Cl_2}, \qquad \mathrm{O=O} \text{ or } \mathrm{O_2}, \qquad \mathrm{N\equiv N} \text{ or } \mathrm{N_2}$$

The noble gases have no valence bonds, and their atoms never pair with one another to form molecules or combine with other elements. Thus, neon is written Ne to indicate that it exists as single atoms. It is never written Ne_2, which would indicate that its atoms were united in pairs to form molecules.

Changing Valence

The valence of certain elements has long been known to change, or to be different, when the element forms certain compounds. For example, sulphur will unite with hydrogen to form hydrogen sulphide (H_2S) and give us the foul-smelling gas that is rightly associated with spoiled eggs. Here sulphur has a valence of 2. However, when sulphur is burned in air, it produces sulphur dioxide (SO_2), the choking gas that is so noticeable when a sulphur candle is lighted. The valence of sulphur in this compound is 4, since oxygen itself has a valence of 2. Under proper conditions sulphur may combine with oxygen to form another compound, sulphur trioxide (SO_3), and thereby produce that most important industrial chemical, sulphuric acid. Here the valence of sulphur is 6. These three compounds are represented graphically as follows:

$$\mathrm{H-S-H} \qquad \mathrm{O=S=O} \qquad \begin{matrix} \mathrm{O} \\ \| \\ \mathrm{O=S=O} \end{matrix}$$

There are a number of other elements which, like sulphur, are found to have different valences when they form different compounds.

It might be thought at first that certain elements having different valences would complicate the situation to the point where it was hopeless to try to decipher or understand it. It did make for plenty of trouble when the earlier chemists were attempting to reduce all these details to order and to discover the fundamental laws that governed such behavior. However, once the underlying causes were discovered, they could be reduced to relative simplicity. This is often true regarding a fundamental process of nature. Multiple valences now help in understanding just what it is about the elements that determines valence; and, also, they help to make clear what the relationship of the manner in which elements form compounds is to the number of electrons in the outer shells of the atoms.

Let us consider again the two top horizontal rows of the Periodic Table, each of which consists of eight elements grouped as given below. It should be remembered at the same time that hydrogen has one extranuclear electron. This electron is in the first shell, which leaves this shell short one electron and gives to hydrogen a valence of 1.

Group number	0	I	II	III	IV	V	VI	VII
Element	He Ne	Li Na	Be Mg	B Al	C Si	N P	O S	F Cl

Group 0, as shown, contains the two elements helium and neon, both of which are inactive gases. The helium atom has two extranuclear electrons in the first shell, making it complete, or full; therefore, it has no shortage of electrons in its shell. In other words, there are no valence electrons, and the atom is inactive. The same is true of neon, with ten electrons. Two of these are in the first shell, and eight are in the second so that both shells are complete. This vertical column in the Periodic Table is labeled Group 0, meaning that its elements have zero valence. The next group, labeled Group I, has in it the elements lithium and sodium. Lithium has three extranuclear electrons,

with two in the inner and one in the outer shell. Under it is sodium, with eleven extranuclear electrons. Two of these are in the first shell; eight in the second; and again one in the outer shell. This one electron in the outer shell is the valence electron and gives to these elements one valence bond. The column is therefore called Group I.

Similar reasoning will show that beryllium and magnesium have two electrons each in their outer shell. These elements have a valence of 2, and the column is labeled Group II. The elements under Group III each have three electrons in their outer shell and therefore have a valence of 3. Likewise, those under Group IV have four electrons in their outer shells and thereby have a valence of 4, as has previously been noted for carbon. Now, when we come to Group V, we begin to find elements that have two valences. Each of these elements will have five electrons in the outer shell. This makes these shells short three electrons of being complete, since eight is the number necessary to complete the shell. The elements will have a valence of 5 when they enter into compounds in which they give up or lend their valence electrons to other elements. However, they will have a valence of 3 when they enter into compounds in which they take up or borrow electrons from other elements, as three electrons are the number necessary to complete their outer shells.

The elements of Group VI may have more than one valence for the same reason. When these elements combine with hydrogen, they do so by borrowing electrons to fill their outer shell. Since they are short two electrons, they require two hydrogen atoms and, therefore, have a valence of 2. This accounts for sulphur's having a valence of 2 in hydrogen sulphide, as previously mentioned, and for oxygen's having a valence of 2 in water. But when these elements combine with oxygen or similar-acting elements, they lend their outer-shell electrons to the other atoms. They have six such electrons to dispose of, and their normal valence in this respect will be 6. Thus, the valence of sulphur being 6, as noted in sulphur trioxide, is explained. Sulphur also has a valence of 4 in the compound sulphur dioxide. In this compound it apparently releases only four of the electrons in its outer shell.

Group VII consists of elements that also have more than one valence. The atoms of these elements are short one electron in their outer shells; and when they combine with hydrogen they do so by borrowing its one electron. In forming these compounds the valence is 1. However, when they unite with other elements, they do so by lending their outer-shell electrons, and then the valence will be 7 or a smaller number.

With this in mind it is possible to summarize somewhat as follows the valence of the different groups of elements in terms of the electrons in their outer shells:

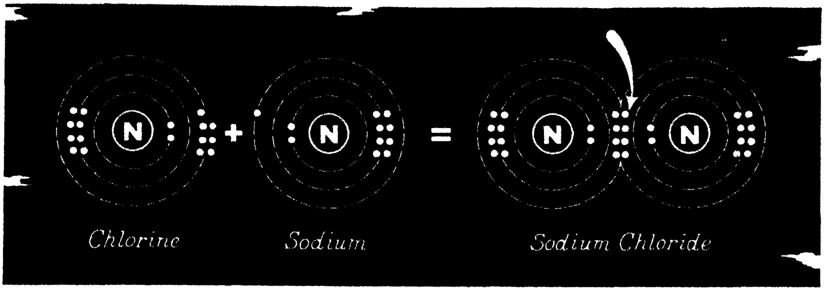
Group number	0	I	II	III	IV	V	VI	VII
Valence to hydrogen	0	1	2	3	4	3	2	1
Valence to oxygen	0	1	2	3	4	5	6	7

The valence bonds are now seen to be the number of electrons in the outer shell of an atom that it may share jointly with another atom when the two elements combine to form a compound. The particular number of such electrons that may be shared and the exact manner in which they are shared determine exactly how the elements behave when they enter into chemical reactions.

Borrowing and Lending Electrons

Borrowing as a human trait may often develop into a strong and unsavory habit, but borrowing and lending electrons between atoms is a common occurrence in nature. It is one of the methods previously mentioned by which atoms transfer electrons and thereby are able to form compounds. A little consideration here may serve to make clear how a great many chemical changes that affect our lives continuously take place by this process. The other method of electron transfer is by atoms mutually sharing electrons, and this will be considered in the following chapter.

One of the most common and widely known chemical compounds is sodium chloride, or table salt. The two elements sodium and chlorine unite and are held together by chlorine borrowing an electron from sodium. Of the seventeen extranuclear electrons in the chlorine atom, two are in the first shell, eight are in the next, and seven are, therefore, in the outer



Graphic representation of electron transfer from sodium to chlorine in chemical reaction to form sodium chloride.

shell. This makes its outer shell one short of being complete and gives chlorine a valence of 1. Sodium has eleven extra-nuclear electrons, two of which are in the first shell and eight of which are in the second. The other electron occupies a position in the third, or outer, shell of the atom and thereby gives sodium a valence of 1. The two atoms will readily combine, with the chlorine atom borrowing one electron from the sodium to complete its outer shell. The sodium by lending one electron also has what then becomes its outer shell complete. This borrowed electron provides the energy that holds the elements together in the compound NaCl, as shown in the diagram.

The information that chemists secured enabling us to know that the aforementioned type of reaction takes place by the transfer of an electron has been acquired in a number of ways and by the use of various mathematical deductions. It is possible to show in a practical way, too, that the force holding elements together in such compounds is electrical in character.

Iron and oxygen are two elements that will combine readily by electron transfers from the iron to the oxygen. One of the well-known results of this reaction is ordinary rust. The two elements may be brought together under special conditions in such a way that, although the atoms may unite, the iron cannot transmit its electrons to oxygen by direct contact; but it may do so by sending the electrons through wires, since wires of certain metals conduct the electrons easily. The wires can be connected to a sensitive electrical meter which will read any flow of electric current through them. In case the chemical activity of iron toward oxygen is brought about by the iron giving up electrons

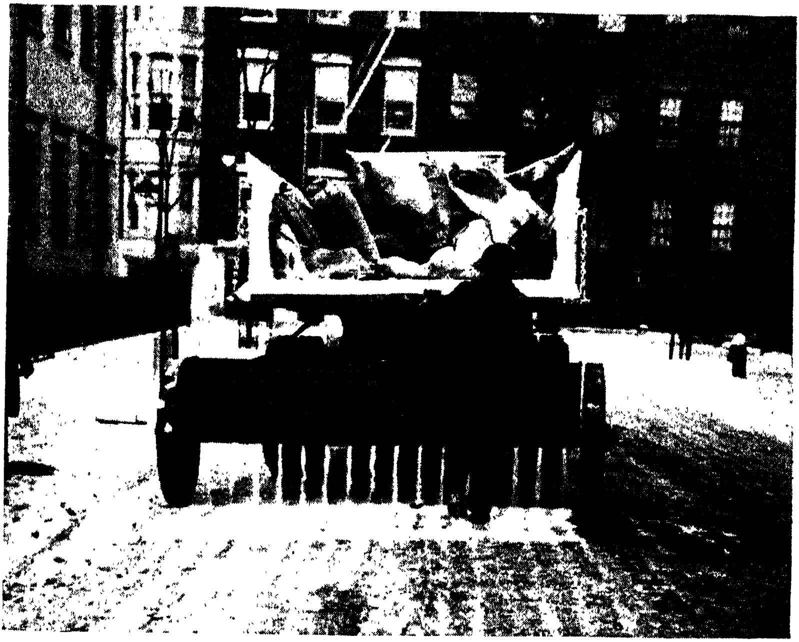
and the oxygen capturing them, the electrons will flow from the iron through the wires and the electric meter to the oxygen, and there will be a reading of the meter. In such an experiment it is found that iron oxide is formed and that at the same time the electric meter shows a reading. This proves that oxygen borrowed electrons from the iron in the reaction and indicates the electrical character of chemical reactions.

Ion Action

The losing or gaining of electrons by atoms in the illustrations given above occurs also in many other instances. Should the atom of any element lose an outer electron or gain an extra one, its electrical balance would be destroyed and it would become electrically charged. When the atom of sodium loses an electron, the atom itself then has one excess positive charge, the positive charges, it should be remembered, being in the nucleus of the atom. Likewise, when the atom of chlorine gains an electron, the atom has one excess negative charge. Such electrically charged atoms are known as ions. The ions may be easily and clearly expressed by use of the chemical symbols. For example, the positive sodium ion is represented as Na^+, and the negative chlorine ion is written Cl^-.

A large number of compounds form ions of their elements when the compounds are in water solutions. These charged particles become dissociated in the solution and move about in random fashion throughout the liquid. Such solutions are called electrolytes, as was mentioned at the beginning of this chapter. Electrolytes have very interesting and peculiar properties, all of which are based upon the losing and acquiring of electrons by their atoms.

One of the properties of all solutions is that when a compound is dissolved in water, the freezing temperature of the solution is lowered and the boiling temperature raised. Should a little sugar be dissolved in water and the solution cooled until ice crystals begin to form, the temperature will be lower than that of the freezing point of water. Just how much lower the freezing temperature is will be determined by the number of molecules of sugar dissolved in the solution. That is, the amount the freezing point of the solution is lowered is directly pro-

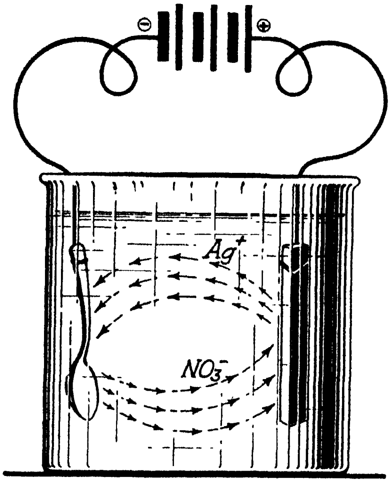


Melting snow on the streets of New York City by sprinkling calcium chloride on the snow. The salt lowers the freezing temperature of the snow-salt solution below the temperature of the air, thus causing the snow to melt. (Ewing Galloway photograph.)

portional to the number of particles in solution. This is true not only of a solution of sugar in water but of any substance dissolved in any liquid. It is this principle of lowering the freezing temperature of solutions that is used in making antifreeze liquids for automobiles. These antifreeze substances almost invariably consist of a mixture of a liquid that will take on a finely divided form in another liquid. The freezing temperature of the mixture will be determined, in part, by the number of suspended or dissolved particles present per unit volume. By increasing the concentration of the mixture, the freezing temperature may be lowered to several degrees below zero Fahrenheit, as most motorists know.

In the case of electrolytes, however, it is found that the freezing temperatures are lowered more and the boiling temperatures are raised higher for a given amount of the compound dissolved in water than is true for the same amount of a nonelectrolyte. For example, dissolving a given amount of salt in water will lower its freezing temperature more than will dis-

solving a proportional amount of sugar in it. Anyone who has ever frozen ice cream in a home freezer knows that salt rather than sugar is mixed with the ice. This would be a more economical process even if the price of the two commodities were proportionately the same. The explanation of the salt's behavior in this respect is that the salt molecule in going into solution ionizes into two separate particles, Na^+ and Cl^-, thereby doubling the number of particles present in the solution. Thus the greater degree to which electrolytes have their freezing points lowered and their boiling points raised is accounted for by the additional particles present in the form of ions.



Ion action in silver plating a spoon from a bar of silver.

Another characteristic of electrolytes is that they will conduct an electric current. An electric current will not pass through pure water; however, when salt is dissolved in it, the solution becomes a good conductor. Even soap in water makes it a fair conductor of electricity, as will be attested to by persons who have received an electric shock while taking a bath and handling a faulty electrical appliance at the same time. The ionized particles in solutions seem to act as little "boats" to carry the units of the electric current through the solution.

Furthermore, when an electric current is passed through an ionized solution, the substance in solution is decomposed. This reaction is the basis of much of the electrochemical industry as well as electric-battery action and electroplating. Silver nitrate ($AgNO_3$), for example, when put into solution is ionized to form silver ions (Ag^+) and the nitrate radical ions (NO_3^-); and when an electric current is passed through the solution, the current carries the silver ions to one of the electrodes. Here the silver ions take on electrons from the current to replace those lost in ionization and form metallic silver atoms (Ag). The silver is deposited on the electrode, where it builds up a spongy coating

of silver if the process is continued long enough. If a brass spoon is substituted as the electrode to receive the silver, it will eventually become covered with silver. The process of silver being "plated out" on the receiving electrode will continue until the silver nitrate in the solution is exhausted. If it is desired that the process be made continuous, more silver must be added to the solution in order to form additional silver ions to operate with the nitrate ions left free by the removal of the silver. This is accomplished by making the other electrode a bar of silver. Thus, the metal goes into solution at one electrode and plates out at the other.

In commercial silver plating, silver nitrate is not used, since the process gives an unsatisfactory spongy layer of silver over the surface and may even build up soft, treelike fingers around the coated object. To eliminate this difficulty another silver compound, usually a complex silver cyanide, is dissolved in water to form the electroplating solution. The principle, however, is the same, and it may be illustrated with the simpler silver nitrate. Now, if the reader is willing to exercise a little mental effort, let him follow this reaction and see what happens.

First the silver nitrate ionizes when dissolved in water; *i.e.*, the compound temporarily separates into two units, Ag^+ and NO_3^-, each of which has an electric charge, as indicated below.

$$\underset{\text{silver nitrate}}{AgNO_3} \rightarrow \underset{\text{ionization}}{Ag^+ + NO_3^-}$$

Then the silver ions (Ag^+) migrate to the spoon that is connected to the negative pole of the battery. The negative charge on the spoon attracts the positive silver ions to it, and each ion picks up an electron from the current. Having secured this extra electron, the silver ion then becomes a normal metallic silver atom and sticks on to the spoon.

$$\underset{\text{silver ion}}{Ag^+} \quad \text{plus an electron} \quad \rightarrow \underset{\text{metallic silver}}{Ag}$$

Now, a different process is taking place back at the bar of metallic silver connected to the positive pole of the battery. Some of the atoms of silver give up to the current an extranuclear electron, since the negative charge of the electron is attracted by the positive charge of the battery. The atoms that

are so affected become positively charged and form silver ions, as indicated below.

$$\text{Ag} \quad \text{minus an electron} \quad \rightarrow \quad \text{Ag}^+$$

These silver ions immediately break away from the silver bar and enter the solution to replace those which have been plated out on the spoon, and the reaction goes on and on.

Many other types of reaction occur with electrolytes, in addition to that explained above where one of the elements is removed from the solution. In some cases, one element may displace another in a compound and thereby produce another compound. For example, the metal zinc will displace hydrogen from most acids in reaction. A typical reaction is represented as follows:

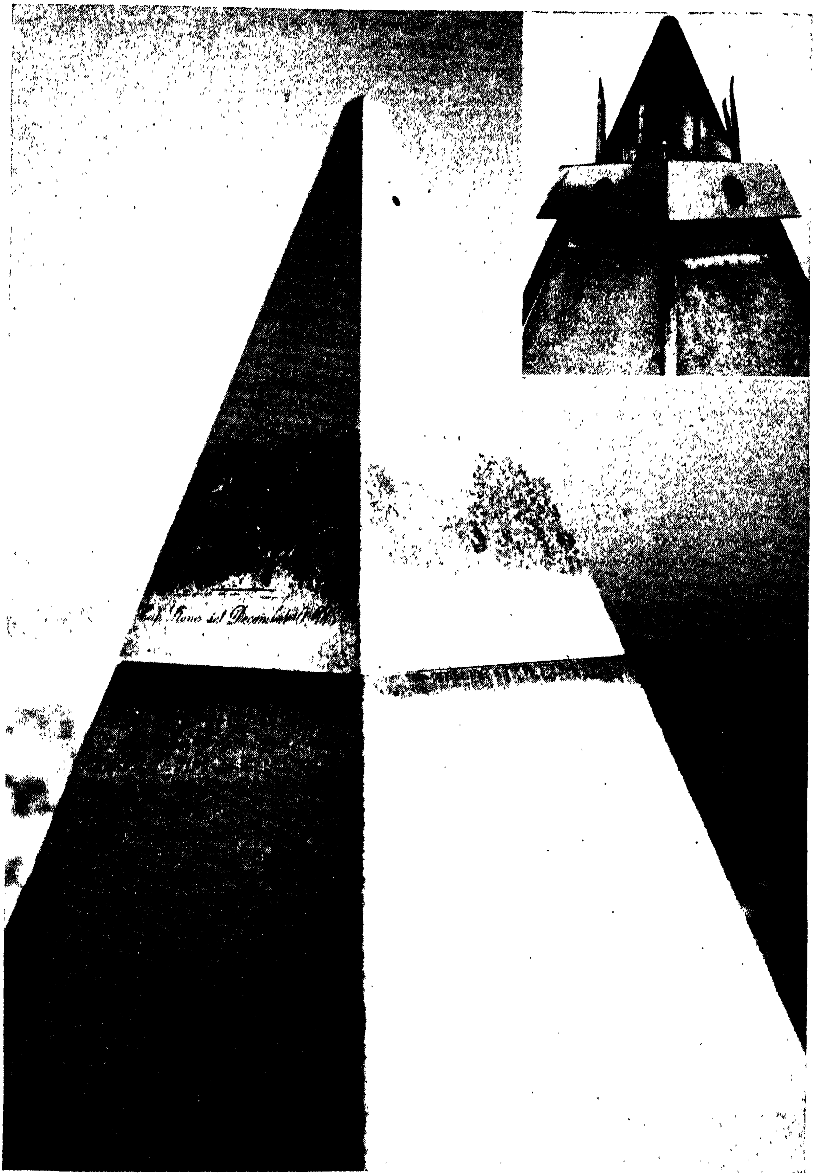
$$\underset{\text{zinc}}{\text{Zn}} + \underset{\substack{\text{sulphuric}\\\text{acid}}}{\text{H}_2^{++}\text{SO}_4^{--}} \rightarrow \underset{\substack{\text{zinc}\\\text{sulphate}}}{\text{Zn}^{++}\text{SO}_4^{--}} + \underset{\text{hydrogen}}{\text{H}_2}$$

Zinc shows a greater tendency to form ions than does hydrogen to remain as ions, and zinc atoms soon lose two electrons when the metal is placed in the solution. As soon as the zinc ion appears in the solution, the hydrogen ions become completely dissociated from the sulphate ion (SO_4^{--}). Each hydrogen ion so liberated soon gathers up an extra electron from its surroundings, becomes a normal hydrogen atom, and escapes as hydrogen gas from the conflict going on in the solution. At the same time, the zinc ion becomes more closely associated with the sulphate ion to form the compound zinc sulphate.

The replacement of one element in a compound by another element that forms more active ions is responsible for many chemical processes. The element potassium, for example, forms active ions and may be used to replace many metals from their compounds that are difficult to liberate otherwise. Before the discovery of the present electrochemical industrial method of preparing aluminum metal, it was prepared in the laboratory by the action of potassium on aluminum salts. The reaction is as follows:

$$\underset{\text{potassium}}{3\text{K}} + \underset{\substack{\text{aluminum}\\\text{chloride}}}{\text{Al}^{+++}\text{Cl}_3^{---}} \rightarrow \underset{\text{aluminum}}{\text{Al}} + \underset{\substack{\text{potassium}\\\text{chloride}}}{3\text{K}^+\text{Cl}^-}$$

The chemistry here is a straightforward displacement reaction in which the aluminum atoms, having regained their



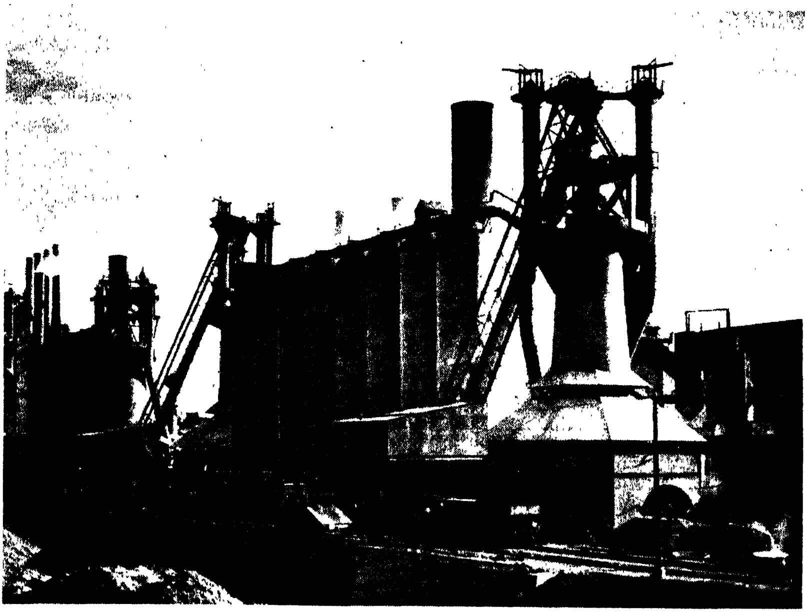
The cast aluminum tip on the Washington Monument as it appeared in 1934 after a period of fifty years. No weathering had occurred, and the engraving was still clearly legible. The band of surface corrosion was produced by galvanic action of copper salts deposited on the aluminum from the bronze collar used to support the lightning rods placed on top of the Monument at the time of its construction in 1884, as shown in the insert photograph. (Courtesy of Aluminum Company of America.)

electrons, become normal metal atoms and are precipitated out of the solution as metallic aluminum. This is a slow and expensive way of preparing it, and in earlier days when it was the only known method of preparing the pure metal, aluminum was more expensive than gold. When the Washington Monument was finished in 1884 there was serious debate as to whether it would be more appropriate to cap the top point with aluminum or gold. The discussion finally reached the halls of Congress where the merits of both gold and aluminum were presented. Those most concerned were convinced that aluminum was more enduring than gold, and so the more precious (at that time) aluminum was decided upon. It seemed to symbolize best the respect of both builders and lawmakers for their distinguished countryman. Today a small pyramid of aluminum rests on top of the monument, not only signifying the expressed desire more than half a century ago to cap this great stone structure with the most enduring metal but also itself giving evidence of the remarkably enduring qualities of aluminum.

Metal Lenders

It must be evident from the foregoing discussion that the metals behave characteristically in the way in which they form compounds. They do so by the atoms lending electrons to form positive ions. A few other elements have the same property. Because of this behavior, elements are sometimes divided into two groups: the metals and the nonmetals. Elements showing the property of losing electrons are classified as metals, although not all of them have the other properties usually associated with metals; and those which accept or borrow electrons in reactions are classified as nonmetals.

In a chemical reaction some metals readily lend their free, or valence, electrons to other atoms, and it has been found that the common metals, such as zinc, iron, copper, and potassium, are of this type. These metals enter into chemical reactions easily, and one very common and significant practical result is that they rust and corrode. The atoms of the precious metals, silver, gold, and platinum, give up their electrons less easily, and the same is true to a certain extent of aluminum. They are, therefore, far more stable in that they do not form compounds



Exterior view of blast furnace at the right with gas ovens for supplying heated air at the left. The incline to the top of the furnace has tracks for carrying ores and raw materials to the furnace. The molten metallic iron is drained off the bottom of the furnace, as shown at the lower right. A closer view of molten metallic iron being drawn from a blast furnace is shown in the photograph at the beginning of the chapter. (Courtesy of U. S. Steel Corporation.)

readily. Gold, platinum, and aluminum, for example, do not rust or corrode, and they are some of the most enduring free elements.

Iron is probably the most important metal to man because of certain properties. It is easily obtained from the ores found in nature, and it is plentiful. Iron is present in small quantities in most soils and rocks, which gives them their reddish color. It is assimilated by both plants and animals and is fundamental to life processes. The daily routine of the average person in a great city takes him from home in a steel-frame apartment, to ride a steel subway train on a steel track to an office or factory in a steel-frame building, and probably to work with instruments of iron or steel. Such experiences are matched in some form or other all over the nation.

Iron has been used as a metal for over four thousand years. The first metallic iron used by man probably came from meteorites found on the earth's surface, since meteorites are mostly

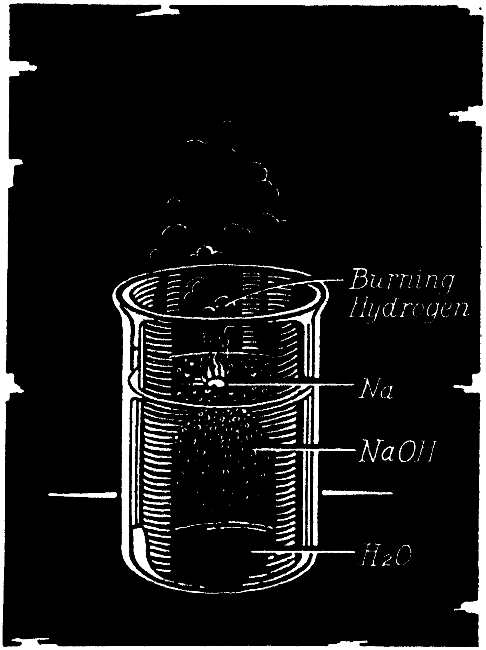
metallic iron. Early Babylonian documents report agents of Babylonian firms working in the "backwoods" of Asia Minor and relate finds of iron by "prospectors," who evidently earned their livelihood by looking for meteoric scraps. Commercial iron is not chemically pure but has mixed in it traces of other elements, chiefly carbon, which give it hardness and strength. The ores from which most commercial iron comes are mainly the oxides, such as hematite, Fe_2O_3; limonite, $(Fe_2O_3)_2 3H_2O$; and magnetite, Fe_3O_4. In some respects an interesting ore is pyrite, FeS_2, which is a brass-yellow, crystalline mineral resembling gold, that has often been mistaken by prospectors for the more precious metal. The mistake has occurred so frequently that the ore is popularly referred to as "fool's gold."

The metallurgy of iron is accomplished mainly by causing carbon to unite with the oxygen, forming carbon dioxide and free iron in a large tower called a blast furnace. The process makes an impure iron called cast iron, which contains about 4 per cent of carbon. Steels of many kinds are made by adding other metals or carbon to the cast iron after the carbon contents of the cast iron has been reduced to about 0.2 per cent. Usually these foreign elements form tiny crystals, or fibers, which mass together and make the iron stronger.

One important property of metals is that they form a group of compounds known as hydroxides. Probably the most distinctive feature of hydroxides is that they contain one or more OH radicals and the soluble ones form negative hydroxide ions, OH^-. The general properties of these soluble hydroxides are that they have a somewhat bitter taste, they produce certain color changes such as turning red litmus blue, and the hydroxide ions will combine with positive hydrogen ions to form water. The most active of the metallic hydroxides are those formed by the alkali metals, *viz.*, lithium, sodium, potassium, and rubidium. Sodium hydroxide may be taken as an example to show a few of the properties of soluble hydroxides. This hydroxide is easily formed by placing some sodium in water, which produces a violent reaction as the following chemical change occurs:

$$\underset{\textbf{sodium}}{Na} + \underset{\textbf{water}}{H_2O} \rightarrow \underset{\substack{\textbf{sodium}\\\textbf{hydroxide}}}{Na^+OH^-} + \underset{\substack{\textbf{hydrogen}\\\textbf{gas}}}{H}$$

In this reaction one free electron from the sodium atom actively displaces a hydrogen bond to the oxygen atom in water and so attaches the sodium atom to the OH radical, one hydrogen atom being liberated. The sodium, by losing one of its electrons to oxygen, becomes positively charged and produces the sodium ion (Na^+). Likewise, the OH group, by receiving one electron from the sodium, becomes negatively charged and forms the hydroxyl ion (OH^-). When the remaining water is evaporated, dry sodium hydroxide as a white solid is secured.



Reaction between metallic sodium and water.

One of the fundamental properties of sodium hydroxide is that it will combine with acids to form water and the salts of those acids. For example, sodium hydroxide and sulphuric acid react to form water and sodium sulphate. Sodium hydroxide also reacts with many fatty acids to form fatty-acid salts, better known as soaps. Something like a hundred thousand tons of sodium hydroxide is used annually in the United States for making different kinds of soaps. Most of the hard soaps are made from sodium hydroxide and beef suet. A particularly high grade of soap may be made from sodium hydroxide and olive oil which are the chief ingredients of Castile soap. In addition to soapmaking, sodium hydroxide enters into many other industrial processes, with something like a total of eight hundred thousand tons of the hydroxide being used annually in the United States. The most important of these other uses are in the manufacture of rayon, the paper industry, refining petroleum, and the manufacture of other chemicals.

Glass made invisible by process discovered by Dr. Katherine Blodgett of General Electric Research Laboratory. This glass has an exceedingly high transmission for light. The left half of the clock face has over it an ordinary piece of glass while the right side is covered with the invisible glass. A strong light on the clock produces reflection and glare on the ordinary glass while the invisible glass remains quite transparent. (Photograph by Katherine Blodgett.)

The salts of the alkali metals are often referred to as alkalies, and one of the most important from an industrial standpoint is sodium carbonate. Large quantities of this material are made either directly or indirectly from sodium chloride, of which there are extensive natural deposits. Nearly two and a half million tons of sodium carbonate, industrially referred to as soda ash, was produced in the United States in 1940. Its most important uses were in making glass, in the manufacture of other chemicals, in making soap and other cleansers, and in the pulp and paper industries.

Glass is a relatively complex substance which consists of quartz or silicon dioxide, SiO_2, fused with sodium carbonate, Na_2CO_3, and usually with lime, CaO. Ordinary window glass is primarily a sodium and calcium silicate glass, with a chemical composition of approximately $Na_2O{\cdot}CaO{\cdot}6SiO_2$ and slight traces of manganese dioxide (MnO_2). Special glasses are usually made

by the addition of other chemicals to the basic ingredients of glass. For example, Pyrex is made by the addition of 5 to 10 per cent of boric oxide to the silicon dioxide along with some aluminum carbonate. This process produces a glass that will not break with shock or sudden changes of temperature.

Some Nonmetal Substances

One of the many important groups of compounds formed from the nonmetals is the acids. Hydrogen is the one substance common to all acids; therefore, the acid properties in a compound are to be attributed to hydrogen. An acid may be defined as a compound of hydrogen with some nonmetal or nonmetallic radical, and it may be identified by certain properties. Chief among these properties are the following: its aqueous solution has a sour taste; it produces certain color changes, such as turning blue litmus red; it dissolves certain metals and will react with hydroxides to form water and salts.

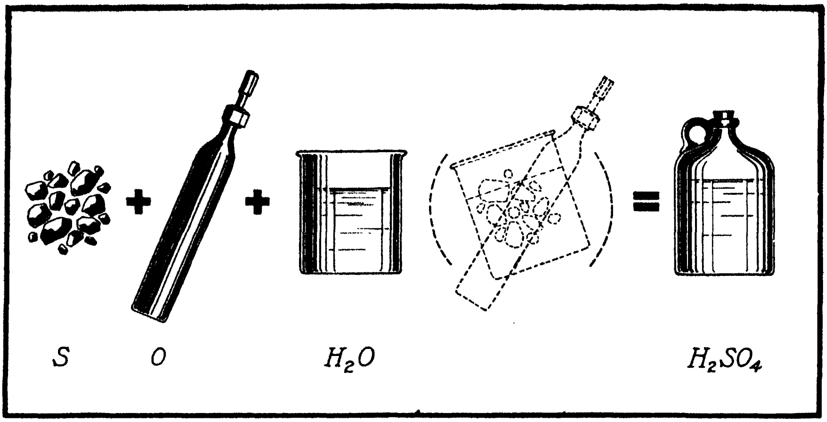
Familiar to most people who have had any direct contact with chemistry is hydrochloric acid. It portrays all the properties associated with typical strong acids. Its water solution ionizes readily to form positive hydrogen ions, and its formula in solution is properly expressed as H^+Cl^-. The measure of strength of any acid is the degree to which the hydrogen ions are formed; and because of possessing this property to a marked degree, hydrochloric is a strong acid. It is employed in industrial processes and in the laboratory in such a great variety of applications that any discussion of its uses is beyond the scope of this text.

In many respects the most important acid is sulphuric, which has the formula H_2SO_4. In solution it ionizes as $H_2{}^{++}SO_4{}^{--}$. Sulphuric acid is manufactured on a large scale from sulphur, oxygen, and water; also, it is a by-product of a number of industrial processes that produce sulphur dioxide as a waste gas. The essential chemical reactions in its manufacture are given in the following equations:

$$\underset{\text{sulphur}}{S} + \underset{\text{oxygen}}{O_2} \rightarrow \underset{\text{sulphur dioxide}}{SO_2}$$

$$2SO_2 + O_2 \rightarrow 2SO_3$$

$$\underset{\text{sulphur trioxide}}{SO_3} + \underset{\text{water}}{H_2O} \rightarrow \underset{\text{sulphuric acid}}{H_2SO_4}$$



Sulphuric acid may be formed from sulphur, oxygen, and water.

Civilization in a country might almost be measured by the amount of sulphuric acid used; at least it is a good indication of the extent to which a civilization has been industrialized. For example, sulphuric acid is used in metal industries to clean metal. It is essential in electric storage batteries, in the manufacture of fertilizers, in what is called the "washing" process of refining petroleum, and in the manufacture of other acids, and it is a common laboratory reagent.

Double Exchange of Chemical Partners

Not all the seething activity in electrolytic solutions results in one element displacing another from its established chemical union. A double exchange of partners often takes place, and this is scientifically referred to as double decomposition. Some of the reactions are very common in the everyday experiences of druggists, chemists, and doctors and in many industrial processes. Nitric acid, for example, has played an important role in history since the ninth century, when it was first prepared by the alchemist Geber. For many centuries it was used extensively to separate gold from silver, and it was employed by Napoleon to prepare more effective explosives than the enemies in his military campaigns possessed. Even now it is in the front rank of acids. Until recent times it was prepared by heating a mixture of sodium nitrate and sulphuric acid in cast-iron stills. The more volatile nitric acid thus formed was allowed

to escape into glass condensers where it was condensed and collected. The reaction was as follows:

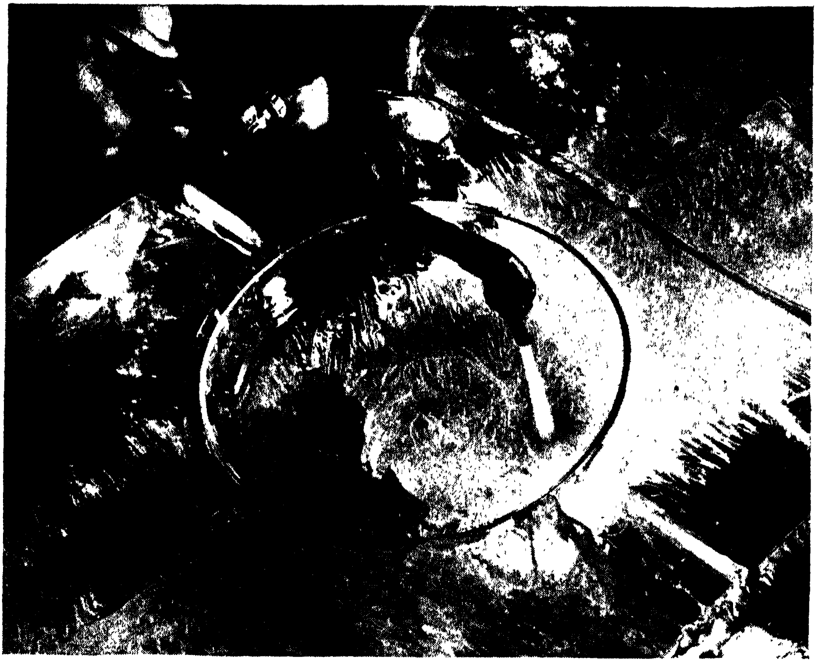
$$\underset{\text{sodium nitrate}}{2NaNO_3} + \underset{\text{sulphuric acid}}{H_2SO_4} \rightarrow \underset{\text{sodium sulphate}}{Na_2SO_4} + \underset{\text{nitric acid}}{2HNO_3}$$

Here the sodium and hydrogen have exchanged compounds, the sodium in the sodium nitrate replacing the hydrogen in the sulphuric acid and the hydrogen combining with the nitrate radical (NO_3). Two new compounds have thereby been formed: sodium sulphate and nitric acid.

Incidentally, this reaction may be used to explain the process of balancing an equation that represents a chemical change. Since nothing is created or destroyed, all atoms present at the beginning must be accounted for after the chemical process takes place; likewise, all atoms in the resulting compounds must be provided before the reaction occurs.

In the preceding reaction, two sodium atoms are required to combine with one sulphate radical (SO_4), as sodium has a valence of 1 and the sulphate radical a valence of 2. However, the sodium nitrate molecule contains only one sodium atom, so two sodium nitrate molecules must be used, as indicated by the numeral 2 in front of $NaNO_3$. In these two molecules there will also be two nitrate radicals. The two atoms of hydrogen released from the sulphate radical unite with the two nitrate radicals. The valence of hydrogen is 1, and that of the nitrate radical is also 1. They unite to form HNO_3; and as two such molecules are formed, they are represented by the numeral 2 in front of HNO_3. The sulphuric acid contains two hydrogen atoms but only one sulphate radical. This is represented by the numeral 2 written at the lower right side of the H, with the numeral 1 understood at the lower right side of the SO_4. Similarly, two atoms of sodium and one sulphate radical make up the compound sodium sulphate, and it is written Na_2SO_4. There are, then, two hydrogen atoms, two sodium atoms, and two nitrate radicals on each side of the equation and one sulphate radical on each side. The equation and the reaction are therefore balanced.

Perhaps the most important reactions of the double-decomposition type occur when a strong acid and a strong base combine



A special type of shrunken glass developed by the Corning Glass Works will stand severe heat changes without breaking. The glass is shown undergoing the blow-torch test while on a block of ice. (Life Magazine photograph.)

to neutralize each other and form water and a salt. By neutralization is meant the process whereby the acid properties associated with the positive hydrogen ion H^+ in the acid and the basic properties associated with a negative hydroxide ion OH^- in the base disappear. These ions disappear from the solution by uniting with each other and neutralizing their electric charges to form water. The real product of the neutralization, then, is water. The other ions remain as they are, forming an attraction for each other to produce a salt.

Should we follow through the process of neutralizing a strong acid, such as hydrochloric acid, with a strong hydroxide, such as sodium hydroxide, we might gain an insight into what occurs. The chemical action may be followed by placing in the solution a few drops of the vegetable dye litmus, which is red in acid, blue in hydroxide, and violet in a neutral solution. The reaction that takes place is as follows:

$$H^+Cl^- \quad + \quad Na^+OH^- \quad \rightarrow \quad H_2O \quad + \quad Na^+Cl^-$$

The hydrogen ions and hydroxide ions have disappeared, forming water. The positive sodium and the negative chlorine ions remain and form the ionized salt, sodium chloride. When the water is evaporated, these ions form a latticework of crystals, familiar to us as white table salt. It might be added here that a good antidote for a strong acid or base, in case a person swallows either, is to take a weak base or a weak acid, accompanied or preceded by large quantities of water. The water will dilute the material swallowed, and the acid or base will neutralize it without serious damage to the patient.

The salts constitute an extremely large group of chemical substances. They are widely distributed in the earth; as one kind or another they are part of all organic life; and they enter into many of our industrial processes. Table salt ($NaCl$), used to flavor foods, is but one of an exceedingly large group of these compounds. Few of them, however, have the peculiar saline taste that we commonly associate with table salt.

REFERENCES FOR MORE EXTENDED READING

HOLBROOK, S. H.: "Iron Brew," The Macmillan Company, New York, 1939.

A highly popularized and personalized story of the development of the steel industry in the United States, emphasizing more its social and economic than its scientific and engineering phases.

SHEER, LEOPOLD: "What Is Steel," translated by F. L. Meyenberg, Chemical Publishing Company of New York, Inc., New York.

Steel discussed in a manner that will be thoroughly enlightening to the interested reader.

HAYNES, WILLIAM: "Chemical Pioneers," D. Van Nostrand Company, Inc., New York, 1939.

Herein is given a vivid account of the founders of the leading chemical industries of the United States.

EDWARDS, J. D., F. C. FRARY, and ZAY JEFFRIES: "The Aluminum Industry," McGraw-Hill Book Company, Inc., 1930.

These research scientists of the Aluminum Company of America have written a thorough and comprehensive story of the technical processes of aluminum metallurgy and manufacture and also included considerable material relating to the development of the aluminum industry.

LUCAS, A.: "Forensic Chemistry and Scientific Criminal Investigation," Longmans, Green & Company, New York, 1935.

An interesting account of the use of chemistry in criminal investigation, treating such topics as blood stains, poisons, fabrics, fingerprints, explosives, and projectiles.

SCHLESINGER, HERMAN I.: "Elementary College Chemistry," Longmans, Green & Company, New York, 1941.

This new text is designed to meet the needs of college students who are not majoring in chemistry or one of its closely related fields. By a careful use of the index, well-written material on the topics discussed in this chapter may be found.

MIALL, STEPHEN, and L. M. MIALL: "Chemistry, Matter and Life," Longmans, Green & Company, New York, 1937.

The first part of this book includes an elementary introduction to many of the fundamentals of inorganic chemistry. The interesting style and excellent organization are likely to cause the reader to continue through the latter part which deals with the simpler phases of the chemistry of living things.

RICHARDSON, L. B., and A. J. SCARLETT, "General College Chemistry," Henry Holt & Company, Inc., New York, 1940, Chaps. XI, XVII, XX–XXIII, XXV.

The chapters noted above include an excellent but not too advanced discussion of the periodic law, chemical families, valence, and ionization.

ENGLISH, G. L.: "Getting Acquainted with Minerals," McGraw-Hill Book Company, Inc., New York, 1934.

A well-known expert in the field of mineralogy has written an engaging story of the study and identification of minerals, using clear diction and a wealth of illustrations.

The Chemical Age, published by Bouverie House, 154 Fleet Street, London.

A weekly journal devoted to industrial and engineering chemistry.

Journal of the American Chemical Society, published monthly by the American Chemical Society, Easton, Pa.

A magazine that contains technical articles on both organic and inorganic chemistry topics.

General Motors.

7: MAN-MADE

Or the Application of Chemistry to Modern Life

ARTHUR JAMES BALFOUR, British scientist, philosopher, and one-time prime minister, observed that chemistry is "a science which pervades the whole of life." It would be presumptuous indeed to imply that the whole range of chemistry affecting modern life is to be treated in these pages. Many complete books have been written on the subject. Some of them are documents of interest and charm in which are portrayed the varied accomplishments of man in creating new substances or adapting natural materials to human needs. To discuss only a few of these achievements here will perhaps give an insight into the processes whereby modern living has been so greatly enriched.

We are said to live in an iron age; at least it is true that iron is at the foundation of our industrial progress. The metallurgy of iron and the manufacture of steel are primarily chemical processes. The same is true of many of the desirable alloys. (Alloys are essentially mixtures of metals or a solution of one metal in another, in very exact proportions and carried out under carefully controlled conditions so as to give desirable and specific qualities. These qualities are usually such things as greater strength, hardness, lightness, malleability, or durability than one of the metals alone possesses.) The swift-moving streamlined trains are a product no less of the chemical laboratory than of engineering factories, since they must of necessity be made of metals that have strength and as little weight as possible. The production of gasoline and other motor fuels is a world industry now outranked possibly only by the iron and steel industry. The refining of motor fuels and oils has come to be largely a chemical process. The production and use of rayon now greatly exceed that of natural silk, and rayon is distinctly a synthetic chemical product of the twentieth century. Furthermore, we depend upon products from American chemical factories to purify our water, preserve our foods, and act as medicines in the prevention and cure of disease.

A Few Underlying Principles

The multitude of chemical processes that affect so much of our living is so great that the nontechnical layman is usually content to use their benefits and to assume that any insight into them is beyond his understanding. This is not necessarily so. The study of a few underlying chemical principles may open up an appreciation for, if not a general understanding of, many processes involved in manufacturing our necessities and conveniences. The purpose here is to consider a few of the general principles and to note briefly how some of them work out in practical applications.

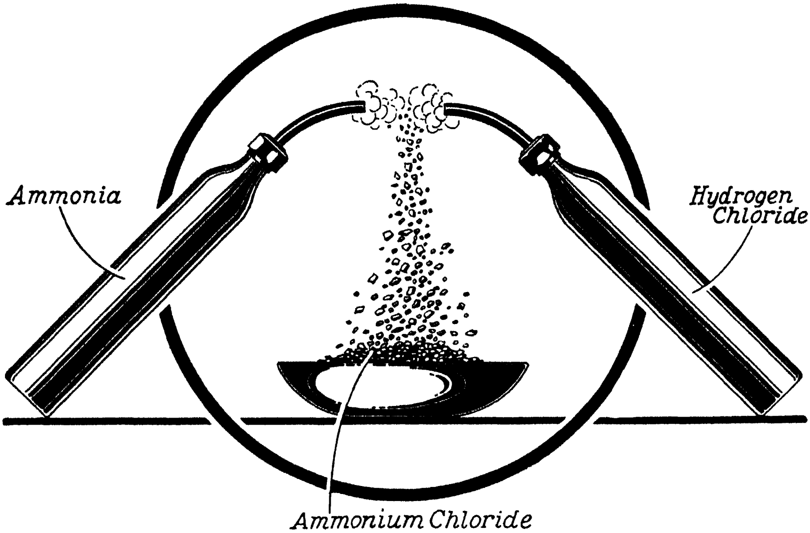
A great many of the articles used in modern civilization are made artificially from raw products or from less desirable refined products. In fact to give the entire list is to recount hundreds of products now enjoyed by man. Such artificially made substances are the result of what is referred to as synthetic

chemistry. Another fundamental process is chemical analysis, by which natural as well as man-made products are taken apart to determine their constituent elements. The purity and fine quality of many commercial articles, drugs, and food supplies are secured by a highly accurate technique of chemical analysis and testing. A third fundamental type of practical chemistry is the metallurgy of natural ores in order to make them usable to man. The metals from which tools are fashioned or great power plants and other structures are built were created in nature for purposes far different from those to which man has put them. The ores from which they come are compounds of the metals, oftentimes mixed with other compounds of little industrial value. They must first be refined. The winning of metals in useful forms from their ores is but one of many instances in which natural substances must be changed or refined before they are satisfactory for human uses. Let us consider these three fundamental processes.

Some Synthetic Substances

Our common usage of the word synthetic as applied to man-made commercial products has given a very general meaning to the term. We speak of synthetic dyes, synthetic silks, synthetic gasolines, synthetic foods, synthetic medicines, synthetic plastics. In general, synthetic substances are thought of as being materials compounded from other products and having quite different properties and characteristics from those of the original ingredients. Obviously a term so broadly used cannot be very exact or scientific; it must naturally apply to a great many processes. Chemical synthesis may be brought about in a number of ways. To be a little more specific, one way is to cause the proper elements or compounds to combine directly and thereby give the desired product. Another is to replace one or more elements by other elements in a compound and in so doing bring about the formation of new and useful products. Another method is to cause an exchange of two or more elements in different compounds. Still another is to break down a complex compound or mixture into simpler substances.

A surprising example of a simple synthesis of two substances combining to form a compound is the formation of ammonium



Synthesis of ammonium chloride "snow."

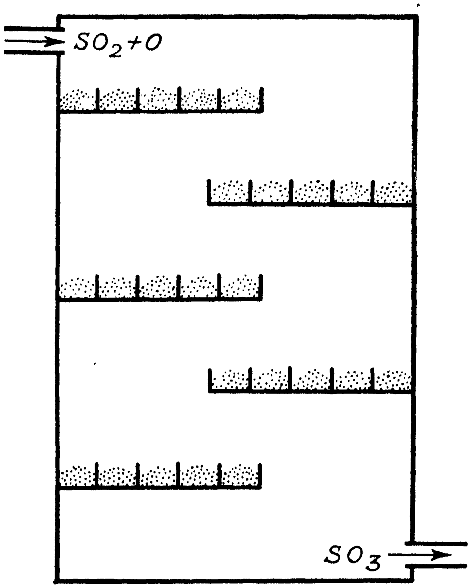
chloride by combining hydrogen chloride with ammonia gas. In this case two invisible gases, when properly brought together, combine rapidly to form a white solid, ammonium chloride, sometimes called ammonium chloride snow. The reaction that takes place is expressed as follows:

$$\underset{\text{ammonia}}{NH_3} + \underset{\text{hydrogen chloride gas}}{HCl} \rightarrow \underset{\text{ammonium chloride}}{NH_4Cl}$$

Likewise, sulphuric acid is made chiefly by the relatively simple type of synthesis whereby a useful compound is built up from its constituent elements or from simpler compounds. Sulphur is caused to combine directly with oxygen to form sulphur dioxide, or sulphur dioxide from other industrial operations is used. The sulphur dioxide is made to combine with more oxygen under suitable conditions to form sulphur trioxide, which in turn is combined with water to form the acid. The equations for these chemical reactions were noted on page 203.

In this process the pure sulphur mined from the great deposits of Louisiana and Texas is burned in air to form sulphur dioxide. This reaction proceeds easily when the sulphur is ignited. However, to get the sulphur dioxide to combine with more oxygen to form sulphur trioxide is a difficult task. It is

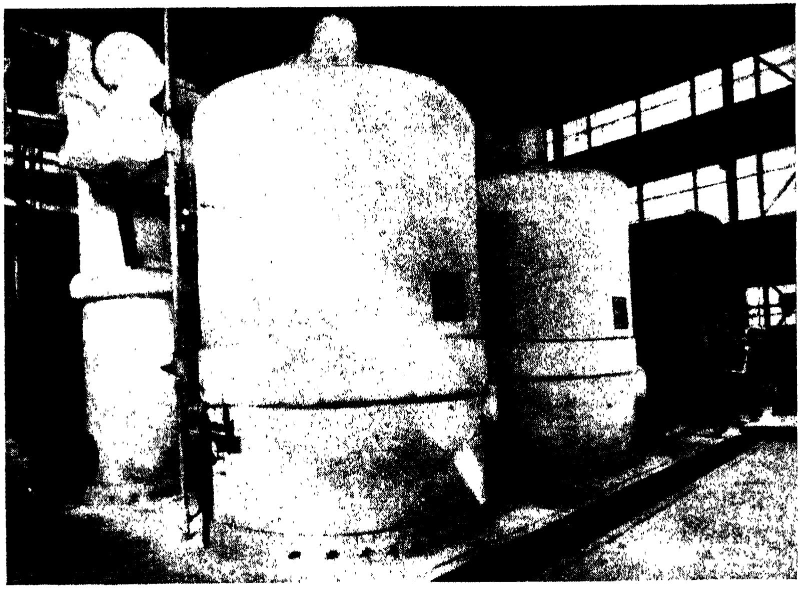
now accomplished in what is known as the contact process. A mixture of pure sulphur dioxide and air that has been thoroughly cleaned of all impurities is led through a series of chambers containing a suitable catalyst. Until recently, the catalyst most extensively used has been finely divided platinum metal.



Sulphur dioxide and air are forced through a converter tower containing shelves on which platinized asbestos facilitates the union of the gases to form sulphur trioxide.

The platinum is distributed over a fibrous material such as asbestos in order to keep it separated and exposed over a large surface. The platinized asbestos is usually laid out on suitable shelves which alternate with each other throughout the chamber. In passing through the chamber, sulphur dioxide and oxygen come in contact with the catalyst on the asbestos. Their union to form sulphur trioxide is then effected. By maintaining a temperature of approximately 400°C. about 98 per cent formation of sulphur trioxide takes place. The next step is to dissolve the sulphur trioxide in concentrated sulphuric acid, about 98 per cent pure, and then add water to form more acid by the water combining with the dissolved sulphur trioxide.

The demand for this great acid, the most important industrial compound made by man, is so enormous that over ten million tons is produced annually in the United States. Platinum (until recently more costly than gold) used as a catalyst is a financial item of major proportions. It can be used indefinitely without loss unless it becomes "poisoned" by the presence of impurities in the combining gases, particularly mercury and arsenic compounds; the slightest traces of these, however, will affect the platinum so that it will not work and has to be re-



Two converters A and B for the manufacture of sulphuric acid. The converters consist of a series of chambers containing the vanadium catalyst. When a preheated mixture of SO_2 and air is led into them, the SO_2 is converted into SO_3. (Courtesy of Chemical Construction Company.)

Sulphuric acid from the absorption towers is placed in glass carboys or large specially lined steel drums for shipment. (Courtesy of Monsanto Chemical Company.)

moved. Recently another catalyst which is even better than platinum has been found. It consists primarily of vanadium oxide, which is cheaper than platinum and is not so sensitive to poisoning. Its use as a catalyst has resulted in the saving of millions of dollars, because the amount used of even the catalyst is large in such an extensive industry. In this case it is seen that a great chemical industry which is quite technical in its details is based upon rather simple fundamental chemical principles.

Synthesis of Some Carbon Compounds

A great many products now made synthetically and widely used by man are produced from carbon compounds. They include not only a great range of commercial dyes and medicines but also such commonly used products as rayon, bakelite, and many other resinous materials. The chemistry of carbon enters also into the refining of all gasolines and motor oils and into the manufacture of much of the illuminating gas now used. In addition, its reactions are vital processes of all living things. A general understanding of a few of the fundamental types of carbon compounds and their properties may give some insight into the basic chemistry of some of their products and into processes of manufacture that were unknown to man a half century ago.

Let us recall, as was pointed out in the preceding chapter, that elements enter into chemical reaction by an exchange of the electrons in the outer shell of the atoms. This exchange is effected in two ways: one by an actual transfer of electrons from one atom to another, the other by a mutual sharing of electrons between atoms. The first method is characteristic of most of the elements in forming the inorganic compounds, as was discussed in the preceding chapter. In this case, ionization of the compounds usually takes place when they are dissolved in water, and a few typical examples were noted. The second method is characteristic of carbon in its compounds, and it may be briefly considered here.

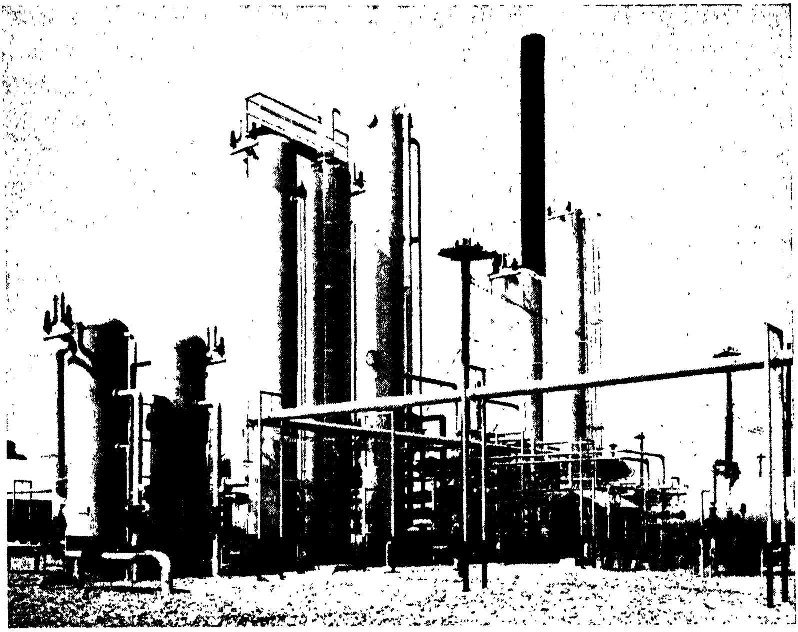
These compounds were formerly classified as organic compounds, since they were built up as a result of organic life and were not found elsewhere in nature. It was believed that a "vital force" took part in their formation in some mysterious

way. However, within the last century man has learned how to produce many of them in the laboratory from inorganic substances, having discovered many of the chemical processes that evidently take place within the bodies of living substances. Today organic compounds are produced synthetically on a wide scale. Some of the main principles that apply to inorganic compounds are known to apply to the organic. As a result, they are more properly referred to as the compounds of carbon.

The compounds of carbon have several characteristics, however, that are not common to the compounds of other elements. One is that very few of them are soluble in water, and those few never form ions. For example, a solution of sugar in water will not conduct an electric current so well as a solution of salt in water. The same is true of other carbon compounds soluble in other liquids than water. This indicates that in carbon compounds there is no actual transfer of electrons from one atom to another but rather a mutual sharing. The shared electrons of the outer shells of the atoms are the bond that holds the atoms together in carbon compounds.

Carbon forms an enormous number of different compounds in comparison to the relatively few formed by each of the other elements. Over one hundred thousand have been thoroughly studied, and the structure of their molecules accurately mapped. An equally large number of others are known but have been studied in less detail. This ability to form such a large number of compounds is accounted for by the fact that carbon is able to share mutually the electrons in the outer shells of its atoms with other elements and with other atoms of carbon; also, by the exact way in which this sharing effects the structure of the resulting molecule. That is, two compounds of carbon may have exactly the same molecular composition, but the atoms are arranged differently in the structure of the molecule, and thereby two different substances are produced, each having properties different from those of the other.

It has been found that when carbon forms compounds, it does so by forming two distinct and fundamental types of structural molecules. One of these is called the chain, the other the ring, type of molecule. The natural oil petroleum is a mixture of compounds of carbon and hydrogen that have



A gasoline refinery where crude oil becomes gasoline. (Science Service photograph.)

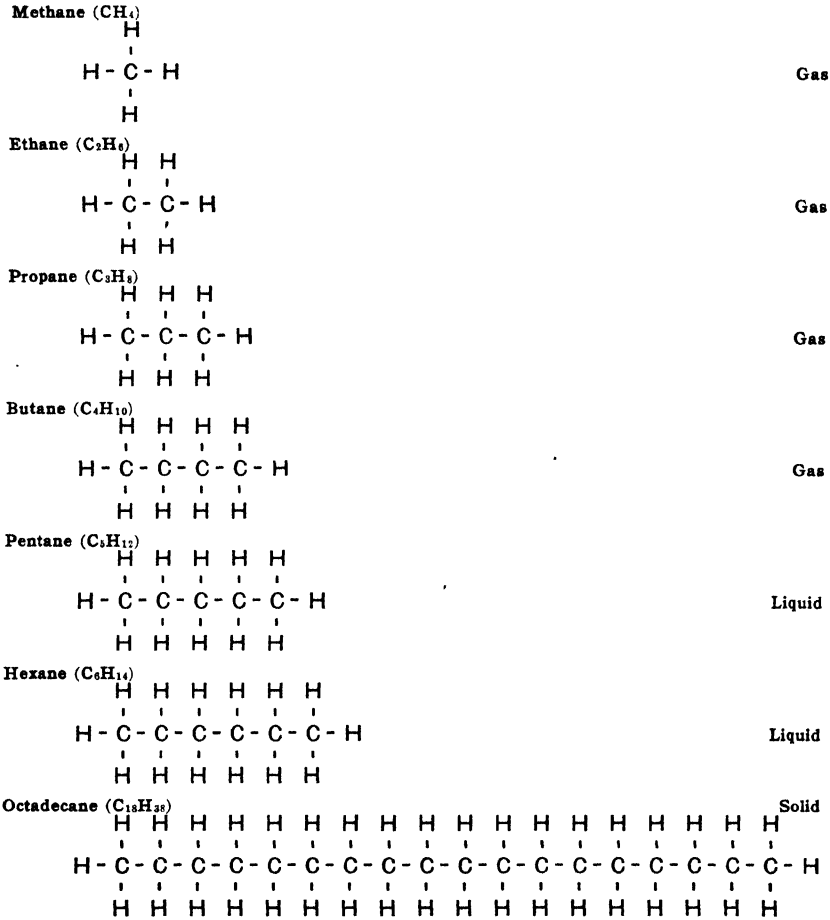
molecules of the chain type. Benzene, a product of coal-tar, is a compound of carbon and hydrogen of the ring type. A brief study of these types of compounds and how it is possible to produce so many synthetic products from them is worthy of our consideration.

Chain Carbon Compounds

One group of the hydrocarbons, as the compounds from petroleum are called, is the methane series. The first member of this series is methane gas, the chief constituent of all illuminating gas. It bears the chemical formula CH_4. The next member is ethane, with the formula C_2H_6; the third is propane, C_3H_8; the fourth butane, C_4H_{10}; the fifth pentane, C_5H_{12}; and the sixth hexane, C_6H_{14}. In order to get an idea of the architecture of the molecules, which affects the properties of these compounds as well as their composition does, it is necessary to look inside them and to represent them by structural formulas. In fact it is always desirable to represent the carbon compounds with

structural formulas, since molecular structure as well as molecular composition is of importance in determining their properties.

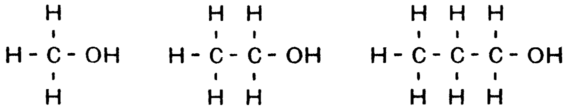
The methane series, in part, may be shown more accurately in diagrammatic form, as is illustrated here.



The one compound above hexane is given to show how the series is continued and that there is a change of properties to a solid as the complexity of the molecule increases. Altogether a great many different known substances are included in the series, the last known one having the formula $C_{60}H_{122}$. Thus, from petroleum are secured various grades of illuminating gas,

gasoline and other liquid fuels, and denser materials such as lubricating oils and solids.

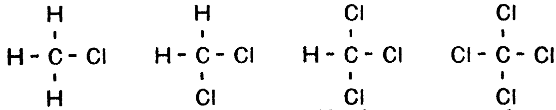
Structural formulas reveal that carbon has a valence of 4 in all these compounds and that its valence electrons may be shared with other carbon atoms and hydrogen. Furthermore, it is possible to substitute other elements for one or more hydrogen atoms, and in each case a different substance is secured. In this manner some of our useful synthetic products are obtained. For example, when an OH radical is substituted for one hydrogen, a series of alcohols is obtained. The first three are methyl (commonly known as wood) alcohol, ethyl (or grain) alcohol, and propyl alcohol, with the following formulas:



wood alcohol — grain alcohol — propyl alcohol

The first of these alcohols is commercially prepared by the destructive distillation of wood, and the second is commercially made by the fermentation of grains; however, by making the proper substitution they may be prepared in the laboratory from the two gases methane and ethane.

Should chlorine be substituted for one or more of the hydrogens, another series of compounds would be obtained. Suppose that we see how this works by considering only the gas methane. Four different substitutions are possible, as follows:



methyl chloride — methylene chloride — chloroform — carbon tetrachloride

Methyl chloride is the gas commonly used in household refrigerators, as it is less objectionable than either ammonia or sulphur dioxide in case of leakage. Chloroform is a well-known anesthetic and is also a good solvent for certain substances. Carbon tetrachloride is a good solvent for fats and, as such, is used as a cleaning fluid; it is also fairly satisfactory as a fire extinguisher. In the aforementioned cases, the theoretically

simple process of substitution permits man to make some of his useful synthetic substances.

The chain hydrocarbons have two very important and characteristic properties which have given us many valuable materials. One is that large chain molecules may be heated under pressure and broken down, or "cracked," into simpler hydrocarbons. The other is that more complex molecules may be built up from the simpler hydrocarbons.

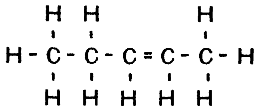
The most widely practiced processes of the first type involve the "cracking" of the hydrocarbons of petroleum by chemical reactions, and this has been of tremendous significance in the petroleum industry. Petroleum consists of a great variety of substances, many of which have complex molecular structures. Some of these are the kerosenes and denser gas oils. For example, the kerosenes are compounds that have from ten to fifteen carbon atoms in the molecule, whereas the gas oils may contain from fifteen to thirty. A typical kerosene is the one having the formula $C_{12}H_{26}$; a common gas oil is one with the formula $C_{20}H_{42}$. The kerosenes and gas oils together constitute about 55 per cent of the petroleum.

There is little or no use in the present automobile industry for the kerosenes as such. The less complex compounds suitable for gasoline constitute a much smaller percentage of petroleum; and if these were the only compounds now used for gasoline, the supply would be much less than it is, and the price would be much higher. The fact that cheap gasoline is available is due to its manufacture by breaking down the more complex molecules of kerosene and gas oils into the simpler ones of gasoline.

One of the kerosenes may be cracked to form an efficient gasoline as follows

$$\underset{\text{kerosene}}{C_{12}H_{26}} \rightarrow \underset{\text{gasoline}}{C_7H_{16}} + \underset{\text{gasoline}}{C_5H_{10}}$$

It will be noticed that one of these products, C_5H_{10}, is short two atoms of hydrogen, as the completed molecule would be C_5H_{12} (see page 218). Such compounds are referred to as unsaturated hydrocarbons, and many thousands of them are now known to the chemical world. The one referred to above might possibly have the structural formula shown on the following page,



in which the two carbon atoms short in hydrogen share an additional pair of valence electrons with each other as indicated by the double bond. Mixtures of such unsaturated compounds in gasolines are very effective in producing antiknock fuels.

This process of chemically manufacturing gasoline from heavier oils by cracking is primarily one of heating these substances to high temperatures under great vapor pressures. Several hundred barrels of gas oils are run into a still that has been built to withstand high internal pressures. Heat is applied until temperatures of about 500°C. are reached; but because of the high vapor pressure on the liquid, boiling does not take place. Pressure in the cracking stills is maintained at approximately one thousand pounds per square inch, which is enough to keep the gas oils from boiling at any temperature reached in the still. The complex molecules, subjected to such enormous forces of heat and pressure, break up into smaller and lighter molecules. Thus, the hydrocarbon $C_{20}H_{42}$ may break up to form two or three molecules containing six to ten carbons each. These lighter molecules are able to boil off under conditions existing in the still and thus pass out into the condenser where they are cooled and collected as liquid gasoline.

By this process of cracking, the production of gasoline from petroleum can be doubled, and even more. The number of gallons of gasoline from a barrel of oil has steadily increased during the last few years, as the cracking process has been improved; even the kerosenes may now be made into gasoline. In 1905 the petroleum industry of the United States had as its main responsibility the supplying of about 33 million barrels of kerosene, most of which was used for lighting purposes, and little use was made of the remainder of the crude oil. The advent of the gasoline engine changed that picture completely. In 1937 the United States produced 571 million barrels of gasoline and 65 million barrels of kerosene. By the most effective cracking process the gasoline yield has been increased to 65 per cent of the volume of crude oil, in contrast to about 20 per cent in a

straight-run distillation. Thus the materials that until twenty-five years ago were the main product of petroleum, *viz.*, kerosene and light oils, are now substantially by-products used primarily to produce gasoline; and gasoline was at that time a dangerous waste which had to be discarded. The chemical process of cracking is, therefore, one of the major factors in oil refining.

In addition, a great host of other substances have been prepared from the materials secured by cracking petroleum. For example, one of the compounds that can be made by cracking the higher hydrocarbons is the light liquid pentane (C_5H_{12}). It, in turn, is used in the manufacture of one of the alcohols that is becoming a valuable commercial solvent. So great are the possibilities of producing different substances in this manner that one corporation has recently spent $10,000,000 in the construction of a plant to produce synthetic chemicals from petroleum. Alcohols, glycols, aldehydes, esters, acids, finil resins, and their derivatives will be among the products manufactured.

The other significant property of the chain hydrocarbons is that more complex molecules may be built up from simpler hydrocarbons. This process has attached to it the imposing name of "polymerization." It is simply the opposite of cracking. Gasoline may be prepared from some of the illuminating gases such as propane (C_3H_8) and butane (C_4H_{10}) and some of the unsaturated hydrocarbons formed as a by-product of petroleum refining. For example, one gasoline has been polymerized as follows:

$$\underset{\text{gas}}{C_3H_8} + \underset{\text{gas}}{C_4H_6} \rightarrow \underset{\text{gasoline}}{C_7H_{14}}$$

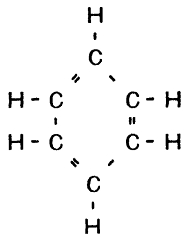
It has been estimated that if the practice of polymerizing the by-product gases that result from petroleum refining were generally followed, two billion gallons of gasoline could be added to our present annual production without any increase in the extraction of petroleum.

Ring Carbon Compounds

In the destructive distillation of coal one of the products obtained is coal tar, a substance that is itself a mixture of several groups of compounds, known commercially as crudes. Crudes may be separated from each other by fractional dis-

tillation, and when so secured they become, in the hands of the chemists, sources of a great variety of valuable substances. Dye colors made from the black, disagreeable, tarry mass exceed in number by far those produced in nature. Modern medicine uses mostly synthetic drugs owing their origin to coal tar; and the perfumes of nature's flowers are much less varied than are the synthetic perfumes produced from this evil-smelling substance.

One of the compounds secured from coal tar that is the beginning of a long series of other substances is benzene, a colorless liquid, lighter than and insoluble in water, and having a boiling point of 80.4°C. The formula is C_6H_6, and the atoms are joined in a ring type of molecule which may be represented as

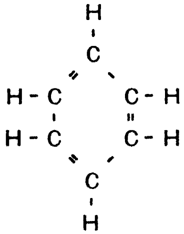


benzene

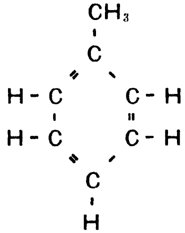
In such a ring structure the carbons are alternately linked to one another by single and double valence bonds. The other valence bonds may be shared by hydrogen, as shown in the foregoing illustration. It may also be noticed from the structural representation of this molecule that carbon has four valence bonds.

Benzene is one of the most versatile materials known to man in the number of compounds that can be made from it by substituting in the ring and by various combinations of rings. For example, when benzene is treated under the proper conditions, one of the hydrogens may be substituted with a CH_3 radical, and the substance known as toluene produced. Toluene is of interest to the public because it is the source of T.N.T. (trinitrotoluene), one of the most highly explosive compounds known to man. Trinitrotoluene is secured by substituting three nitro groups (NO_2) for three of the hydrogens in toluene. The relation of benzene to toluene and to T.N.T. and the types of

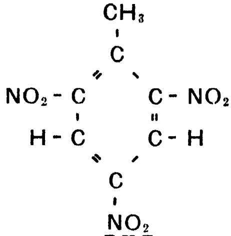
substitutions necessary to produce this explosive are seen by examining their structural formulas.



benzene

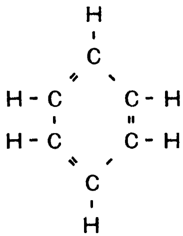


toluene

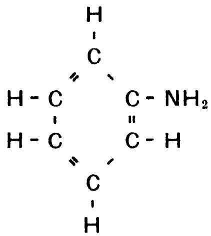


T.N.T.

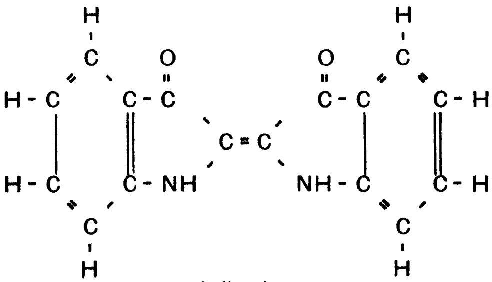
However, when one of the hydrogens of benzene is replaced with an NH_2 radical, another compound, called aniline, is secured. Aniline is, in turn, the first of a long series of compounds that produce the aniline dyes, the number and variety of colors of which have already been referred to. Indigo dye, as an illustration, is a complex combination of a substitution in the aniline derivative of the benzene ring. It is the most extensively used commercial dye, and the United States produces about fifteen thousand tons of it annually, mainly by this synthetic process. The relationship of these three compounds is shown in their structural formulas.



benzene

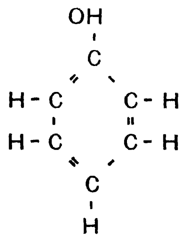


aniline

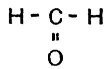


indigo dye

Another large group of synthetic substances made from various carbon compounds are the plastics, so extensively used in modern industrial and commercial life. Plastic products are familiar to us as fountain pens, phonograph records, buttons, decorative jewelry, dental and surgical instruments, electrical switchboards, instrument cases, telephone receivers, armatures and commutators for electric generators and motors, and a variety of other things. Perhaps one of the best known of the plastics is bakelite. One of the main varieties of this, as well as of a number of other trade-name products, is technically known as a phenolic plastic. This class of plastics is so named because one of the chief raw products of its manufacture is phenol, a coal-tar product, commercially called carbolic acid (C_6H_5OH). The other essential ingredient is formaldehyde ($H \cdot CHO$), a substance secured by the oxidation of wood alcohol. Phenol is a hydroxide derivative of the benzene molecule, and the structural formulas of phenol and formaldehyde are as follows:

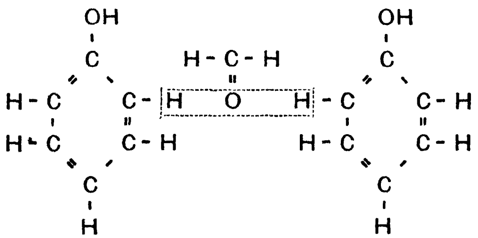


phenol



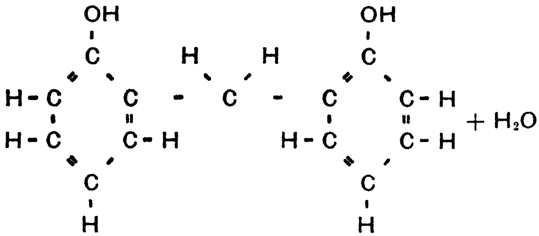
formaldehyde

When the two substances are made to combine, a large number of the ring molecules are linked together by means of the carbon atom of the formaldehyde. The process may be visualized by taking a simplified illustration of two phenol molecules uniting with a formaldehyde molecule.



By eliminating a molecule of water as shown in the dotted rec-

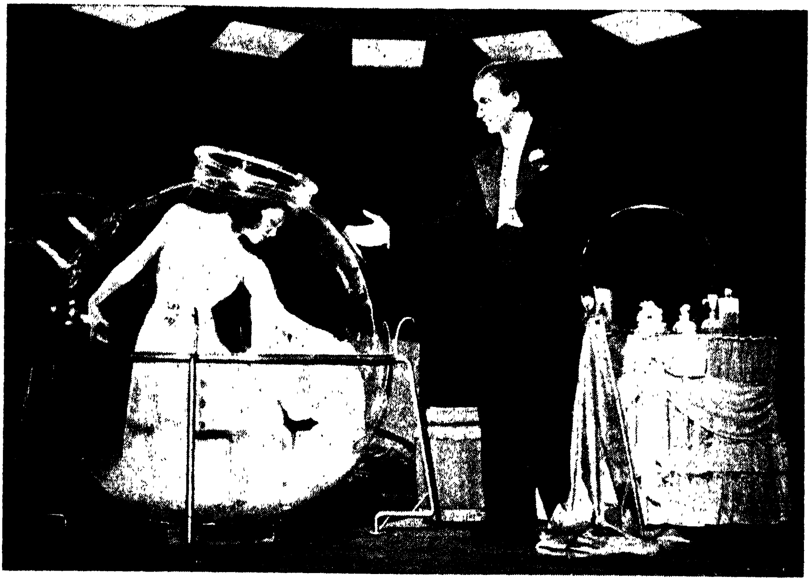
tangle, the double-bond carbon of the formaldehyde forms a connecting link between two phenol molecules to produce the following more complex molecule:



In practice an indefinite number of such combinations can be made to take place, the molecules building up in three dimensions. Such molecules form a noncrystalline and plastic substance; however, one of the most difficult problems was to get the plastic material to set or harden. This was finally achieved by an American chemist, L. H. Baekeland, in 1907, when he reheated the plastic substance under very high pressures and found that under conditions of high temperatures and high pressures it would set into a hard solid which could not be remelted, was insoluble and resistant to acids, would not burn, and was a good electrical insulator. Apparently what happens in the process of reheating under pressure is that the molecules further combine or polymerize, to build up still more complex molecular structures.

Other plastics are produced in a somewhat similar manner from different raw products. Some of them are relatively clear and can be artificially dyed in a wide range of colors. An organic substance known as methyl methacrylate will form a colorless plastic that is just as transparent as glass or more so. It is manufactured and sold under the various trade names of *Lucite*, *Flexiglass*, and *Prexiglass*, the uses of which include lenses, window and automobile panes, even entire automobile bodies, as shown in the photograph at the beginning of the chapter, dental appliances, and various novelties where transparency and nonbreakable qualities are desirable.

It is not possible to describe here in any further detail the interrelations of the compounds of benzene and other coal-tar



Movie actress Loretta Young dancing in a giant retort of Lucite in photoplay "Eternally Yours." (Courtesy of DuPont Company.)

products from which dyes, medicines, perfumes, explosives, and poison gases are derived. The examples just given illustrate the fundamental type of chemical changes that may be accomplished. It will at once be evident that a great many interrelations exist and that through them an enormous number of different compounds may be made. The technical expert knows exactly what the relationships are and how to produce the exact chemical change necessary to secure the desired compound. All that is necessary, therefore, is a supply of the raw material and suitable equipment for its treatment. The available knowledge, the raw materials, and chemical plants have given to America a unique chemical industry and created a wide variety of useful materials.

Synthetic Silk

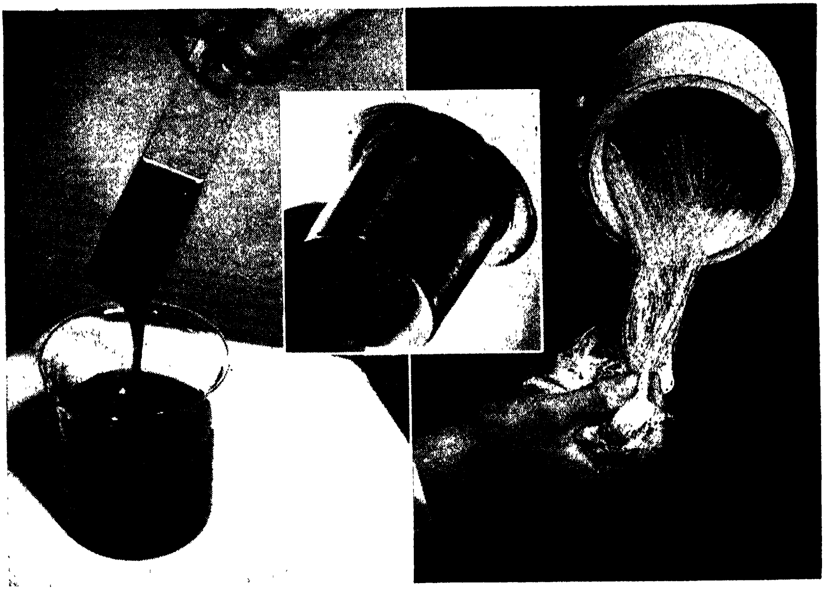
The securing of silk has been one of man's quests since the beginnings of recorded history. We know from reliable sources that silk was used for making fine cloth in China as early as 5,000 years ago. How much earlier than this the Chinese began to use the delicate fibers of the silkworm and to develop seri-

culture is shrouded in the uncertainties of antiquity. Cultivation of the silkworm and the manufacture of silk cloth have constituted an increasing business during recent centuries.

For more than a hundred years man has been attempting to duplicate the processes of the little caterpillar in producing the fiber as it spins its cocoon. Near the close of the last century a commercial process was inaugurated for making artificial silk from wood or cotton fibers, and within the last fifteen years the production of artificial silk, commercially known as rayon, has not only equaled that of natural silk in volume but far exceeded it. In 1912 the United States consumed about 3 million pounds of rayon and 21 million pounds of silk; in 1929 the consumption was 131 million pounds of rayon and 120 million pounds of silk; and in 1938 we used 327 million pounds of rayon and 51 million pounds of silk. Rayon has become one of the major fibers of the textile industry.

Rayon is not made by a process of twisting or spinning the plant fibers together; it is the product of a series of chemical reactions which finally produce a fine, continuous strand of regenerated cellulose. Every step is a chemical process. The initial physical structure of the plant fibers is entirely destroyed, and a new fiber is created that has exactly the qualities desired. Fundamentally, however, the process of making these conversions is fairly simple. It consists essentially in reducing the pure cellulose of plant fibers to a liquid material, spinning the liquid into filaments, and then converting the filament into a modified dehydrated cellulose, the familiar solid filament used in making rayon cloth.

Four methods are used today for making rayon. They are all somewhat similar, but one of them, the viscose process, accounts for about 80 per cent of the world's rayon production at present. As raw material the viscose process uses spruce wood or cotton linters or a mixture of the two. The spruce wood is first treated with calcium bisulphite solution to remove the lignin, a natural gum of wood which sticks the cellulose fibers together. The treated wood and the cotton are then soaked in a solution of sodium hydroxide, which causes the fibers of cellulose to swell and to form a compound known as soda-cellulose. The soda-cellulose is rolled into sheets, after



Photographs illustrating three steps in the making of rayon. Left, liquid viscose; inset, spinneret showing holes through which viscose is forced; right, rayon fibers. (Courtesy of American Viscose Corporation.)

which it is broken into fine crumbs and allowed to age for a definite time at a constant temperature.

The next step is to treat the soda-cellulose crumb with carbon disulphide, a reaction resulting in the formation of yellow-orange crumb, a product called cellulose xanthate. The whole purpose of this treatment is to convert the cellulose to a compound that is soluble in water. When the orange crumb is soaked in water under very carefully controlled conditions, it forms an alkaline, amber-colored, viscous liquid, called viscose. The viscose, after proper aging, is pumped into spinning machines. These are fitted with platinum nozzles containing perfectly machined fine holes. The viscose, in being forced through the minute holes, is divided into very fine streams which immediately pass into a hardening bath. The bath consists primarily of sulphuric acid which neutralizes the alkali of the viscose. Thus the viscose is again converted into pure cellulose which is a solid, and a continuous fiber is thereby produced. The collected filaments are spun into skeins or twisted into yarn for commercial use.

The various methods of rayon manufacture have given man new kinds of fibers with which to work and made possible a remarkable variety of clothing materials. Rayon may be used exclusively in weaving and knitting, or it may be combined with wool, silk, cotton, and linen fibers to produce fabrics of various textures unknown a few years ago. These materials will take dyes in such manner that the designs and colors of cloths are limited only by man's ingenuity in conceiving artistic patterns and in mixing his dyes. Rayon has placed silklike clothing within the reach of all.

Stockings from Coal, Air, and Water

In 1938 it was announced that a new kind of synthetic material had been developed from which fibers rivaling those of finest silk could be made. Christened nylon, it was widely heralded as made from coal, air, and water, and during the following year women's hose of the fabric appeared on the market. Many tests have shown that nylon fibers surpass in strength and elasticity any previously known textile fiber; that they are resistant to fraying or abrasion; that they are not injured by cleaning in boiling water, steam, or dry-cleaning fluids; that they may be made to have a dull luster comparable to that of silk; and that they may be dyed in a wide range of hues. Such a new material not only stimulated the curiosity of the lay public but also was of primary interest in scientific circles.

The discovery and perfection of nylon resulted from an extensive program of research which was designed to investigate the synthetic building up of large complex molecules from smaller and simpler ones. This research finally led to the discovery of how to make a series of chemical substances, scientifically called polyamides, from which it is possible to form fibers, bristles, and sheets characterized by extreme toughness, elasticity, and strength. A polyamide is defined as a man-made protein-like chemical product. This means that it has somewhat the same chemical composition as the proteins, of which silk, hair, and wool are familiar examples. There are many different polyamides, even as there are many different proteins, and it is not possible to assign any specific chemical formula to

Chemistry in modern fashion. Dresses are of rayon, stockings of nylon, shoe heels and hats of plastics. (Courtesy of DuPont Company.)

them. However, in each polyadmide the structure of the molecule is extremely complex, as is also true of the proteins.

One of the simplest methods of making nylon is to build up a polyamide from a dibasic acid and an organic diamine. There are a very large number of dibasic acids that are easily prepared, and an organic diamine is a substance containing two amino radicals which are constituent parts of all proteins. The dibasic acid may be, and usually is, made from coal. In the

process of making a dibasic acid, the coal is destructively distilled to yield coal tar. The coal tar is then fractionally distilled, and one of the products is phenol. Phenol is in turn treated chemically so as to yield the dibasic acid. Air and water serve as sources of nitrogen and hydrogen, respectively, which are a part of the composition of all diamines. These elements are combined under proper conditions of temperature and pressure to produce ammonia (NH_3). Ammonia is easily converted into a simple diamine by replacing two hydrogen atoms with other groups of elements, a typical one of which is dimethylamine, $(CH_3)_2NH$.

The diamines and the dibasic acids each consist of relatively simple chain molecules; but when they are caused to react, they combine to form the complex molecules of a polyamide, from which nylon is made. Molecules of dibasic acid will apparently not combine with each other; neither will the molecules of a diamine combine with other diamine molecules; however, a dibasic molecule will unite with a diamine molecule. This union takes place in a sort of chain fashion in which dibasic acid molecules form alternate links of the chain, and diamine molecules constitute the joining links. By regulating carefully the environmental conditions of the reacting molecules, such as temperature, pressure, and nature of catalyst, it is possible to build up polyamides of long-chain molecules.

The polyamides so formed in the reacting chamber are in a liquid state. The liquid is forced out of the chamber by a suitable pump through tiny holes in a spinneret. As soon as the filaments come in contact with the air outside the spinneret, they instantly cool to a solid state. The size of the filaments may be controlled, therefore, by varying the openings in the spinneret and also by exerting a pulling force on the filament as it solidifies.

One very unique property of nylon is that it can be "cold drawn"; *i.e.*, the fiber may be stretched four to seven times its original length after it has cooled, depending upon the particular polyamide being used. The stretched fiber remains at its new length, and as such it becomes exceedingly strong and elastic. This property of being able to be cold drawn depends upon the

arrangement of the long-chain molecules within the polyamide. The long molecules are arranged at random, much like the individual straws in a haystack. When nylon is cold drawn, the molecules become parallel to one another and are brought much closer together, so that the fiber becomes longer and smaller. The closeness of the molecules to one another adds strength to the fiber and also gives it elasticity so that if it is further stretched, it will return to its cold-drawn length and shape.

The characteristics that nylon acquires when it has been cold drawn, *viz.*, strength and elasticity, are very important; upon them depends its most extensive industrial use at present, that in knitting hosiery. Other synthetically made textile fibers, and also cotton fibers, have sufficient strength, but they lack the elasticity necessary for a stocking to hold its shape after being worn. Silk is the only natural product that has both these properties to the extent required in fine hosiery, but nylon is said to possess them to a greater degree. Many people in the hosiery industry are of the opinion that nylon will soon become one of the most important textile materials for the making of fine stockings.

Chemical Analysis

Analytical chemistry constitutes another important type of practical chemistry. Chemical analysis and testing is a vital part of every major manufacturing process, and government inspection of the purity of drugs and foods is based upon it. It involves exact determination of the elements or compounds in the substances and oftentimes of the amounts or percentages of such elements that are present. There are two kinds of chemical analysis: qualitative, which identifies the elements present in substances; and quantitative, which determines how much of each element or group of elements is present.

An example of the first is a very delicate test for iron. A compound suspected of containing even a trace of iron is dissolved in a weak solution of hydrochloric acid. The acid combines with the iron, if present, to form a chloride of iron. Potassium sulphocyanate is then added to the solution; and should any iron be present, a red-colored substance will be formed. The

reaction that would take place is

$$\underset{\substack{\text{iron}\\\text{chloride}}}{FeCl_3} + \underset{\substack{\text{potassium}\\\text{sulphocyanate}}}{3KCNS} \rightarrow \underset{\substack{\text{iron}\\\text{sulphocyanate}}}{Fe(CNS)_3} + \underset{\substack{\text{potassium}\\\text{chloride}}}{3KCl}$$

The iron sulphocyanate is the only sulphocyanate that is red. Therefore, if the red color appears, it is proof that iron was present in the original compound in order to form the iron sulphocyanate; if no such red color appears, it is positive evidence that no iron was present.

Another example of a qualitative test is one for corrosive sublimate, or bichloride of mercury, a deadly poison. It is soluble in water and alcohol and can be easily put into foods and drinks. As such, it is sometimes used to poison a person whom someone wishes to murder. Any liquid suspected of being poisoned in this manner may be tested in a number of simple ways. One is to put a piece of copper into a sample of it. If any mercuric chloride is present, it will attack the copper, whereas the pure liquid will not. In this event the following reaction occurs:

$$\underset{\text{copper}}{Cu} + \underset{\substack{\text{mercuric}\\\text{chloride}}}{HgCl_2} \rightarrow \underset{\substack{\text{copper}\\\text{chloride}}}{CuCl_2} + \underset{\text{mercury}}{Hg}$$

The mercury is deposited as a black coating on the copper. Nothing else that is soluble in water or alcohol except the poisonous compound bichloride of mercury will give this black deposit on the copper.

The principle that underlies such chemical analysis is to treat the substance being tested in a manner that will reduce the suspected material to a form where it alone will give a definite reaction with a known chemical. When such a reaction occurs, it is clear that the suspected material is present; should it not occur, the material is not present.

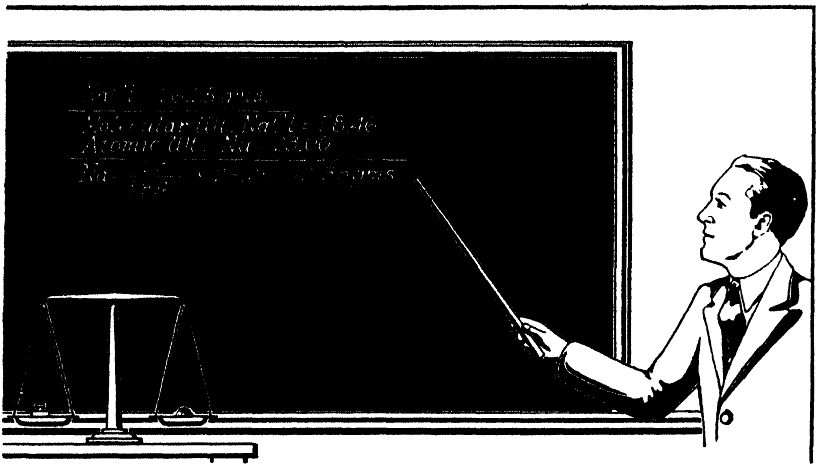
Another standard and exact type of qualitative analysis is made with a spectroscope. By this means a spectrum, or color analysis, of the unknown substance is procured. Each element will give a certain colored light when it is heated to incandescence, and when the light is passed through a spectroscope it is separated into different wave lengths, or spectral lines. This means that each element will emit wave lengths peculiar to that

element alone. Therefore, it is possible by looking at spectral lines of light made by heating any substance to determine just what elements are present in it. This is the method that was used to determine the chemical composition of the sun and other stars, and it is a very convenient and highly accurate method of analyzing any unknown compound for its constituent elements.

In practice a qualitative spectrum analysis is usually carried out by taking a spectrum photograph of the unknown material. Such a photograph may be obtained by using a spectrograph with a camera attachment and by producing an electric spark through a sample of the material placed at the proper position in front of the spectrograph. The material may be illuminated, however, in other ways, such as by putting it on the tip of a carbon electrode or in an illuminated chamber. The spectrum photograph of the unknown material so made is then compared with a standard spectrum photograph, usually of iron, with the places of all the lines of the other elements marked on it. Such a photograph serves as an excellent measuring device. The spectrum lines of the unknown compound may then be matched with the known lines of the different elements, and the unknown thereby identified.

Thus far the only consideration has been to find out whether or not a particular material was present in a given substance. In many cases it is also necessary to determine exactly how much of the various constituents is there. This is quantitative chemical analysis. The practice of quantitative analysis is based upon the fundamental principle of definite composition in chemical compounds. Compounds are made up of individual atoms of the elements, and these atoms have definite and unchanging weights. In analyzing a substance quantitatively various chemical reactions are brought about until the unknown element is secured in a form in which it can be weighed or its volume can be measured. Then mathematical calculations are employed to determine the exact amounts of the elements or compounds that were in the original substance.

For example, in the analysis of compounds for the quantity of metallic salts present, a system has been devised that depends upon the formation of insoluble salts of the metals. In this way



A quantitative chemical analysis of a substance produced 26.25 grams of sodium chloride. From this information it is easy to calculate that 10.33 grams of sodium were present in the substance.

a given metal is precipitated out as a part of the insoluble salt; and after its identity is established, the precipitated compound is accurately weighed. The percentage of metal in the compound may be determined by simple calculation, using the atomic weight values of the elements, and thus the exact amount of the metal present is arrived at.

Again, substances may be analyzed without precipitation and weighing. In this case it is necessary to cause them to react with a solution, the concentration of which is known. This is usually referred to as volumetric analysis. The strength of vinegar, for example, may be determined by treating it with a hydroxide of known concentration. The active agent in vinegar is acetic acid which will, of course, neutralize a hydroxide. Suppose that a 10 per cent concentration of sodium hydroxide is used to react with the vinegar, and twenty-five centimeters of the hydroxide neutralizes 100 cubic centimeters of the vinegar. Since the strength of the hydroxide is known, it is easy to calculate the strength of the vinegar, or the amount of acetic acid in it.

Quantitative analysis requires a high degree of skill and accuracy, but by using the techniques now available it is possible to measure the amount of an unknown present in a

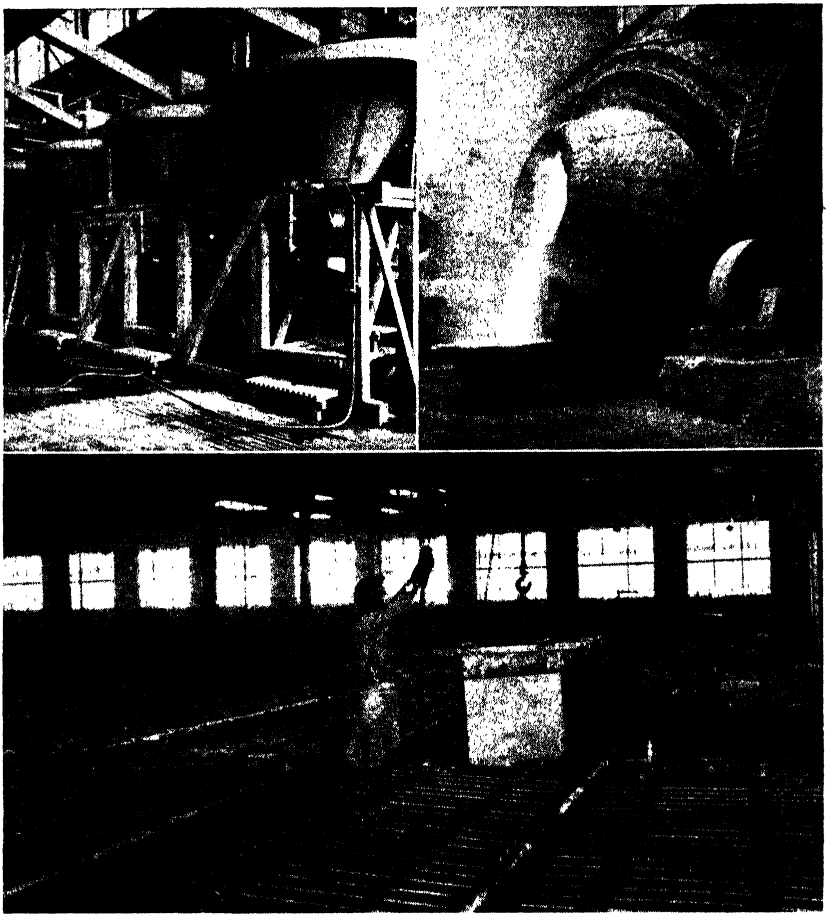
substance in terms of one-thousandths of a gram. This accuracy in quantitative analysis is extremely important in controlling the purity of drugs and in determining the composition of commercial chemical substances. It is such common practice now in the drug and chemical industries that failure of a firm to conform to this exactness means legal indictment in the courts or financial ruin or sometimes both.

Refining Raw Materials

The refining of most natural substances is necessary before they are useful to man. Refining usually consists of the separation, by some kind of chemical reaction, of the desired materials from the raw products found in nature. To mention only a few, native phosphates must be purified before being used as fertilizers; natural petroleum must be refined before it is usable for either gasoline or motor oils; the metals, with only a few exceptions, occur in nature in compounds or ores, which must be refined before any use is made of them.

The metallurgy of copper is an illustration of the manner in which the chemical industry enters into the refining of metals. Let us note in brief detail how it is accomplished and possibly gain from it an insight into the general principles involved in refining raw materials. The world produces over two million tons of copper annually, a small amount of which is found as native metal, but most of which is secured from copper ores. By far the most profitable type of ores are the sulphides of copper, but even these contain 90 to 98 per cent of rock materials not wanted. The securing of commercial copper is, therefore, primarily a metallurgical process and one that involves about five steps.

The first stage is what is known as concentration. Most of the ores contain large quantities of rock in which the copper compound is segregated, and this rock material must first be removed. This is chiefly a mechanical process and is done by a series of crushing, grinding, washing, and flotation processes. Approximately two-thirds of the mined material may thus be discarded, and about 95 per cent of the copper originally contained in the entire mass is concentrated in the remaining third. The other parts of the process are designed to remove the copper



Illustrating three steps in the refining of copper from its ores. Left photograph shows concentration and classification of ores; right picture shows molten metal being poured from a converter; and bottom photograph, electrolytic tank room where 100 per cent pure copper slabs are produced. (Courtesy of Anaconda Copper Mining Company.)

from its compounds and to separate it from all impurities in the ore. They are usually complicated somewhat by the presence of iron sulphides combined with the copper sulphides.

The second step is to roast the concentrated ores in large furnaces which are heated with powdered coal and air. The copper and iron sulphides are separated, and most of the iron is oxidized to form iron oxide, while a considerable part of the sulphur is converted into sulphur dioxide. The chemical reaction

may be represented as follows:

$$\underset{\text{copper-iron sulphides}}{2CuFeS_2} + \underset{\text{oxygen}}{4O_2} \rightarrow \underset{\text{copper sulphide}}{Cu_2S} + \underset{\text{iron oxide}}{2FeO} + \underset{\text{sulphur dioxide}}{3SO_2}$$

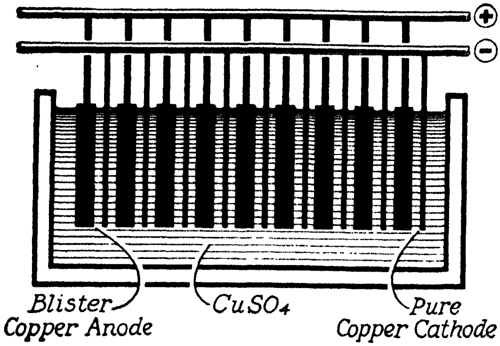
The sulphur dioxide escapes from the furnace as a gas, and some of the impurities, including most of the iron oxide, melt and form a slag which floats on top of the liquid product containing the copper. The slag is removed by pouring it off the copper-bearing liquid. The remaining liquid is about 40 per cent copper, most of which is in the form of copper sulphide. The liquid also contains small quantities of iron sulphide and traces of other metallic compounds, such as sulphides of gold, silver, platinum, and nickel. This molten mass, known as copper matte, is drawn from the furnace to be treated further.

The next step is to place the copper matte in large silica-lined Bessemer converters, each holding about sixty-five tons of charge. An energetic blast of air is sent through the liquid. The remaining iron sulphide is oxidized to iron oxide which readily combines with the silica to form a slag which floats on top. The converter is tilted, and the slag poured off, after which the device is set upright again and a second blast of hot air is admitted. The sulphur in the copper sulphide combines with oxygen in the hot-air blast to form sulphur dioxide according to the following reactions, and it escapes from the converters as a gas:

$$\underset{\text{copper sulphide}}{2Cu_2S} + \underset{\text{oxygen}}{2O_2} \rightarrow \underset{\text{copper}}{4Cu} + \underset{\text{sulphur dioxide}}{2SO_2}$$

The result of this step is to produce molten material that is about 95 per cent pure copper. The copper is further refined in a fourth step by heating it in special refining furnaces where hot air is again passed through the molten material. Any residual iron and sulphur is oxidized. The product obtained from these furnaces is about 99 per cent pure copper and is known as blister copper, since the escape of sulphur dioxide during the cooling process gives the surface of the metal the appearance of blisters. The blister copper is then cast into slabs.

It might be thought that any refining process that produced a 99 per cent pure metal would give a product satisfactory for industrial uses. This is not true, however, with copper. Even



Electrolytic refining of copper.

small traces of impurities make it undesirable for most commercial uses, and it must be further refined. As little as one-hundredth of one per cent of phosphorus in the copper lowers its electrical conductivity 20 per cent; therefore it must be removed, since the largest use of copper is in the electrical industry. The final stage in the metallurgy is to refine the blister copper by an electrolysis process.

In this step, huge slabs of blister copper, weighing about 600 pounds each, constitute the anodes of the electroplating system. These are immersed in a bath of the proper chemical composition. The other electrode consists of a thin sheet of pure copper. When the current is applied, copper atoms from the impure anode migrate to the thin sheet cathode much the same as do silver ions in the silver-plating process, the details of which were previously discussed. After a time all the copper has been transferred, building up the cathode into a large slab of 100 per cent pure copper. If the electrolysis voltage is kept at the exact proper value, the other metals in the blister-copper slab fall to the bottom of the tank as sludge or are dissolved in the electrolytic bath. They may then be removed from the sludge or bath and are themselves often of considerable value. In most copper-refining plants enough platinum, silver, and gold is recovered to pay for the electrolytic process. The amount of silver obtained as a by-product from copper refining is more than is secured from any other source in the United States.

The total amount of copper that has been used in the United States is of the order of several hundred million tons. Here is,

indeed, large-scale chemistry. To discuss the metallurgy of iron and aluminum, and the refining of petroleum and rubber, is to recount a similar story. Each one requires a special process, of course. Usually the fundamental chemistry involved is rather simple; but the methods of bringing about the reactions may be exceedingly difficult and technical.

The short discussion given in this chapter may indicate the multitude of ways in which the science of chemistry enters into every phase of modern life. The attempt here has been not to glorify chemical research and chemical engineering but rather to develop an understanding of the fundamental idea that the chemical industries are based upon well-known and clearly established natural laws and that they do not work in a mysterious and magical fashion.

REFERENCES FOR MORE EXTENDED READING

FINLAY, ALEXANDER: "Chemistry in the Service of Man," 5th ed., Longmans, Green & Company, New York, 1939.

This book is an excellent account of what the science of chemistry has done for mankind. No advanced chemical knowledge is necessary for understanding it.

GREGORY, SIR RICHARD: "Discovery: Or The Spirit and Service of Science," The Macmillan Company, New York, 1931.

The contribution of science to society is told in writing of great beauty and presented from a philosophical viewpoint.

HATCHER, W. H.: "An Introduction to Chemical Science," John Wiley & Sons, Inc., New York, 1940, Parts II and IV.

The sections referred to give in an abbreviated and well-organized manner the fundamental chemistry of organic compounds and some of our great industrial processes.

RICHARDSON, L. B., and A. J. SCARLETT: "General College Chemistry," Henry Holt & Company, Inc., New York, 1940, Chaps. XXX, XXXI, XXXII, XLVIII.

A general understanding of the chemistry of carbon and a few of its compounds may be obtained from the chapters referred to.

MANN, F. G., and B. C. SAUNDERS: "Introduction to Practical Organic Chemistry," Longmans, Green & Company, New York, 1939.

This is a good elementary and brief survey of the fundamentals of the science, presumably for people who will not take advanced work; the emphasis, however, is upon making a technician of the reader rather than presenting organic chemistry from a cultural standpoint.

"The Story of Rayon," The Viscose Company, New York, 1937.

Here is a short description of rayon and its manufacture by the viscose process presented in a popularized style and with many appropriate illustrations.

SIMONDS, H. R.: "Industrial Plastics," Pitman Publishing Corporation, New York, 1939.

It is difficult to keep a book up to date on a subject that is developing as rapidly as the plastics industry; however, the author has produced a thorough and concise study of plastics from the standpoint of their types, properties, processes of manufacture, and uses.

Modern Plastics, published by Breskin Publishing Corporation, Easton, Pa.

A monthly journal containing articles of interest to the layman and dealer in plastics. Extensively illustrated with photographs.

Chemical and Metallurgical Engineering, published by McGraw-Hill Publishing Company, Inc., New York.

Popularly written articles, many of which deal with synthesis and factory processes and contain abundant statistics and graphs.


Ewing Galloway.

8: HEAT AND COLD

A Study of Molecular Motion and Heat

WHEN the American Revolution broke out, there lived in Boston a young military officer named Benjamin Thompson, whose hobby was scientific studies. Unlike some of his contemporaries, he deemed it prudent to go to Europe during the Revolution, and eventually he became a military engineer for the Bavarian government. While serving in this capacity he received the title of "Count Rumford," the name by which he is generally known in the scientific world.

He did not let his official duties interfere with his scientific interests, and in some instances he was able to combine the two. For example, during the process of boring out a cannon barrel he observed that a great deal of heat was produced by the friction of the boring tool. This aroused his scientific curiosity, because he could not understand where the heat was coming from. The

scientists of that time believed that heat was some kind of a tangible fluid, called "caloric," which flowed from warm objects to cooler ones as water flows downhill. But here was a case where no hot objects were present, and yet a vast amount of heat was being created. What was its source? He began searching for the answer.

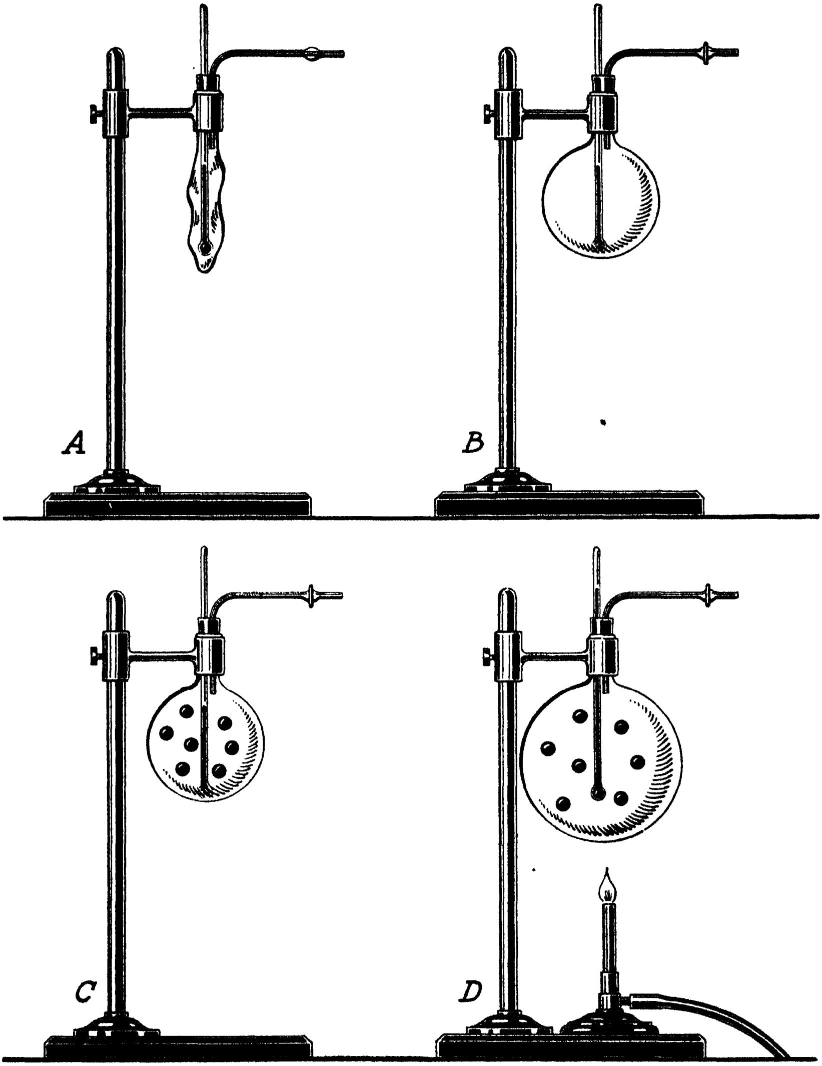
The caloric theory had never been tested by adequate experiments; it had been accepted merely because it furnished a plausible explanation for some of the facts about heat. Count Rumford realized that the theory might be all wrong, so he devised an experiment to check it. The cannon barrel and boring tool were immersed in a tank of water. After the boring process had been carried on under water for a few hours, the water actually began to boil; when the boring motion was stopped, it cooled off. Therefore, the heat was apparently coming from the motion of the tool rather than from the cannon barrel or the tool itself.

From this experiment Count Rumford concluded that heat is simply a form of motion, and this conclusion was presented in a paper before the Royal Society of England. It was considered so revolutionary at the time that few of his colleagues even took it seriously. Not until about half a century later, after it had been confirmed by several other experimenters, was it universally accepted. Today the concept of heat as a form of motion is one of the basic principles of physical science.

What Makes a Substance Hot or Cold?

In previous chapters we learned that all substances are made up of molecules. Although these molecules may vary in size and chemical composition in different substances, they all have some properties in common. One is that they move very rapidly in a random manner. Another is that their speed is increased when heat is applied to the substance. This property is the essence of what makes substances hot or cold. The theory that deals with the relationships between molecular motion and heat energy is known as the "kinetic theory" of heat. We shall now illustrate some principles of it by discussing the molecular behavior of a volume of air enclosed within a toy balloon.

In the accompanying diagram we have a balloon mounted on a laboratory stand and fitted with an inlet tube and a ther-



Deflated balloon (A) is expanded (B) by molecular bombardment of air inside. Molecules shown (C) magnified ten million times, and molecular bombardment is increased (D) by heat.

mometer. It can be inflated through the tube, and the thermometer will register the temperature of the enclosed volume of air. Let us assume that the rubber walls of the balloon are thin enough to be partially transparent. Then, when it is inflated, it will look something like the drawing shown at *B*. The only

thing visible within the balloon is the part of the thermometer that protrudes below the stopper. The rest of the space looks empty because air is an invisible gas.

What is it that keeps the balloon wall stretched out in the shape of a sphere? The obvious answer is air pressure; but just why does air exert a pressure in this manner? The answer to this question involves a study of the molecular behavior within the enclosed volume of air. A balloon six inches in diameter contains about 10^{23} molecules; and since there is considerable empty space between these molecules, it is evident that they must be very tiny. In fact these molecules would have to be magnified about ten million times to be as large as the black dots, shown at *C*, inside the balloon.

Furthermore, all these molecules are in constant motion. The motion is random and haphazard, and it might be compared to that of a swarm of very active bees. The speed of any one molecule may vary greatly from moment to moment, but the average speed for all of them remains constant for any given temperature of the air. For example, if the thermometer is reading 70°F., the average molecular speed is about 1,700 feet per second, a speed equal to 1,150 miles per hour!

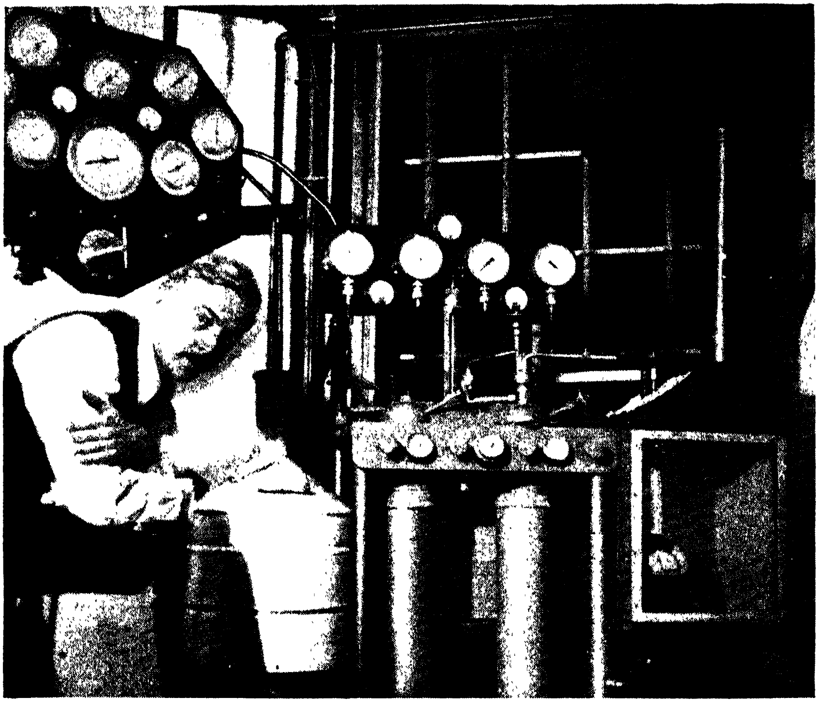
At this great speed no molecule can travel very far without colliding with some of its neighbors. The average distance of travel without a collision is about three-millionths of an inch. This is known as the "mean free path" of the molecule; and since the mean free path is so short, about six billion collisions occur every second. That is why the molecular bombardment of the walls of the balloon appears to be a steady pressure. It might be thought that these collisions would cause the molecules to slow down and stop so that the balloon would collapse. To comprehend the fact that this does not occur necessitates further understanding of the nature of molecular motion.

Molecules have sometimes been compared to tennis balls, but there is one important difference. If we drop a tennis ball on the floor, it will not bounce quite so high as it was originally. In fact it is easy to see that this is impossible; for if it could bounce to its original height just once, there would be nothing to prevent it from doing so a million times, and it could keep on bouncing indefinitely. We know that this does not happen, and

the reason is that a tennis ball is not perfectly "elastic." By this is meant that a small part of its energy of motion is lost each time that it strikes the floor. Molecules do not lose energy, the way a tennis ball loses it, when they collide with one another. Either they rebound with the same original speed, or, if one is slowed down, the other is speeded up by a corresponding amount. The average speed of a group of molecules tends to remain constant; therefore, molecules behave like perfectly elastic bodies: And thus we have one reason why they keep moving in spite of collisions.

If the balloon is heated gently, as shown at *D*, two effects will be noticeable. The size of the balloon will increase somewhat, owing to the expansion of the heated air, and the thermometer will show an increase in temperature. If we could see the molecules themselves, we should observe that their average speed had increased as the heat was applied. It is this increase in molecular speed that is responsible for both the increase in balloon size and the increase in the thermometer reading. If, for example, the thermometer reading increased from its previous value of 70 to 100°F., we should find that the average molecular speed had increased from 1,700 to over 1,800 feet per second. In fact if we made a careful study of the speed of these molecules over a wide range of temperatures, we should find that the average molecular speed and temperature were definitely related. It is beyond the scope of this book to discuss the mathematics of the relation, but it should be kept in mind that any increase in the temperature of a body will mean, in general, that the speed of its molecules must increase.

The same rule applies to the cooling of a body. If, for any reason, the average molecular speed of a given body of matter is decreased, the temperature of that body will drop. This is demonstrated by the fact that a warm object is cooled by contact with an object of lower temperature. The molecules of the two objects exchange impacts over the area of contact, and eventually they all reach the same average speed, a speed that will be slower than the original speed possessed by the molecules of the warm body but faster than the molecular speed in the cooler one. Therefore, the warm body is cooled, and the cooler one is warmed.



Apparatus designed for liquifying hydrogen at a temperature of 269 degrees below zero centigrade. (Science Service photograph.)

The query may have occurred to some whether or not it is possible to cool an object to the point where its molecular motion ceases altogether. In theory this is possible, but such a condition has never been produced in the laboratory. Scientists have calculated that all molecular motion would cease at a temperature of 273.1 degrees below zero centigrade, or 495.5 degrees below zero Fahrenheit. This is theoretically the lowest temperature that could possibly exist, and it is called "absolute zero." By evaporating liquid helium under carefully controlled laboratory conditions, it has been possible to approach absolute zero to within less than 1°C., but it never has been reached. Until it is possible to produce in an object a temperature of absolute zero, the molecules will never cease moving. Should that temperature ever be reached in a substance, all its molecular motion would stop. Under such conditions it is likely that the material will take on unusual properties.

The elementary principles of the kinetic theory of heat which we have just discussed furnish the basis for an explanation of all the facts about heat that we observe in everyday life. The whole theory may be summed up by saying that heat energy in matter and molecular motion are one and the same thing. This means that heat and cold are merely different aspects of the same physical process. This fact is often expressed in the statement that "cold is the absence of heat."

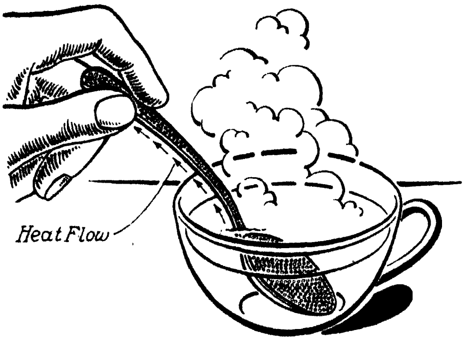
How Does Heat Energy Flow through Matter?

Familiar to most people is the fact that when one end of a metal rod is held in a flame, the other end soon becomes too hot to hold. Apparently the heat of the flame has a tendency to flow through the rod. Similarly, when a silver spoon is used to stir a cup of hot coffee, the handle of the spoon becomes noticeably warm. These and many other simple experiments show that heat flows through solids. The process is called conduction, and the direction of flow is always away from the region of higher to the one of lower temperature.

It is fairly easy to explain the conduction of heat in terms of molecular motion. In the case of the spoon, for example, we can imagine that its internal structure is something like that shown in the drawing on the following page. The molecules in the bowl of the spoon are moving at great speed because of their contact with the hot coffee. Some of them are continually bumping into those in the lower part of the handle. Eventually a part of the increased motion of the molecules in the bowl is communicated, with gradually decreasing intensity, to all parts of the handle. Hence, the apparent flow of heat through the spoon is really a flow of increased molecular motion. The same analysis applies to the conduction of heat through any solid.

Since some solids are much better conductors than others, it is customary to classify solids into two general groups from the standpoint of heat conduction. Those which are quite effective in transmitting heat are classed as conductors, whereas those which are very ineffective are called insulators. This property of heat conductivity is related to the molecular structure of the substance. Some substances will transmit an increase in molecular speed much more readily than others. Most metals

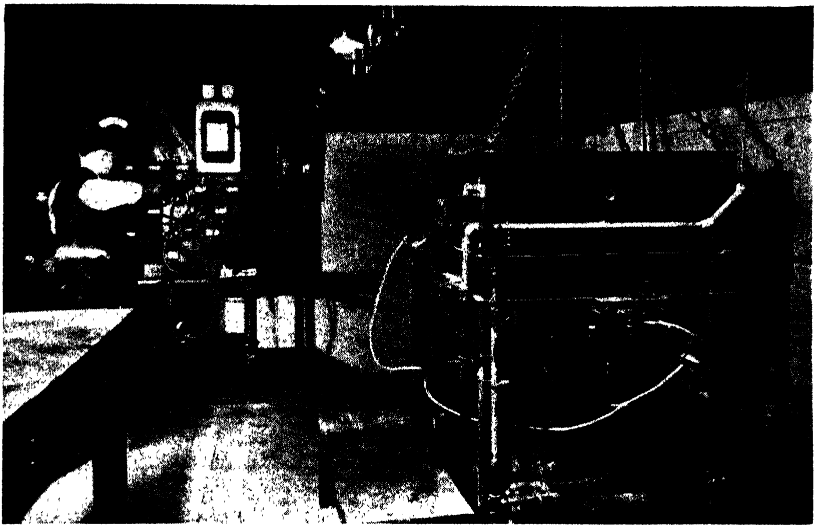
are fairly good conductors; silver, copper, and aluminum are among the best. Some of the most effective insulators are asbestos fiber, diatomic earth, paper, and most fabrics. Copper is about five thousand times more effective than asbestos as a conductor of heat.



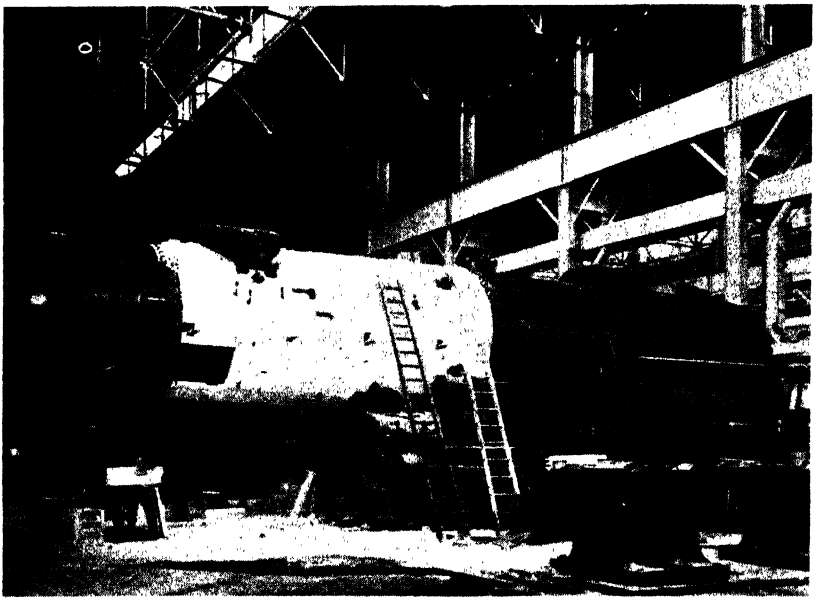
Molecular structure accounts for heat conduction.

Both the properties of conduction and those of insulation have many useful applications; the latter are perhaps more familiar to the layman. For example, the steam pipes in all heating systems are covered with an insulating material, usually consisting of a mixture of asbestos fiber and magnesium carbonate. This is to prevent the loss of heat while the steam is being transferred from the furnace boiler to the radiators. Similarly, all refrigerators are built with double walls and the space between them is filled with a silica compound shredded into a form resembling wool. Such a structure is impervious to the flow of heat and prevents leakage of heat from the room into the refrigerator. A similar type of so-called "mineral wool" is used extensively in the insulation of houses to keep the interior cool in the summertime and prevent the escape of valuable heat in the winter. In this case the material is formed into large sheets for application between the outer and inner walls of the house. It is also prepared in a shredded form which can be blown into the space between the walls, so that insulation may be applied to houses originally built without it.

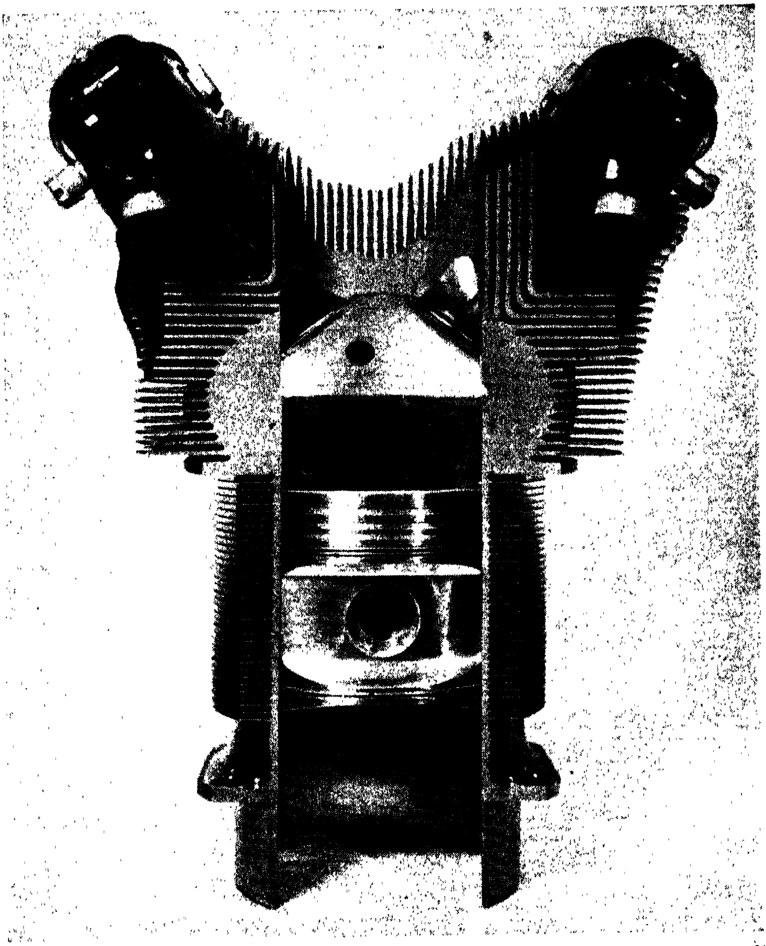
One common example of an application of heat conduction is in the cooling of airplane engines. The cylinders and engine parts associated with the combustion process are surrounded with a large number of radial plates, or fins, of aluminum. This type of structure permits a rapid conduction of heat away from the regions where excessive heat is generated by combustion, and the air removes the heat from the fins. Another practical example of the use of the principle of heat conduction is in a



This is an apparatus for measuring the heat conductivity of materials. In the boxlike structure shown in the foreground, heat is applied electrically to one side of a sample under test. The rate of heat transfer through this sample is then measured by special electric instruments. (Courtesy of Johns-Manville.)



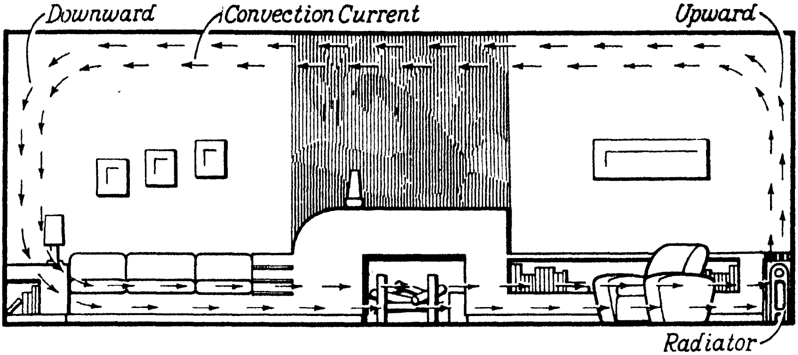
Insulating material shown in white being applied around a steam locomotive boiler to prevent loss of heat. (Courtesy of Johns-Manville.)



The excess heat generated within the combustion chambers and cylinders of an airplane engine is rapidly conducted to the surrounding air by a structure of metal plates or fins as shown in this photograph. (Courtesy of Wright Aeronautical Corporation.)

refrigerating system. The refrigerant is circulated in copper or aluminum tubing so that heat is transmitted easily through the metal tubing from refrigerator to refrigerant and, at another point in the system, from refrigerant to the outside air.

Most liquids and gases are poor conductors, but, owing to their molecular structure, they transmit heat by another method known as convection. One of the best examples of heat transfer in this manner is house heating. When a layer of air immediately



Convection currents flow upward at source of heat and downward at points remote from it.

adjacent to a steam radiator becomes heated, it expands; the expansion decreases its density, and it is therefore appreciably lighter than the cooler surrounding air. Accordingly, the warm air rises to the ceiling, and a new layer of cool air flows in around the radiator to take its place. This new volume of air goes through a similar process, and the lighter air at the ceiling is pushed over to the opposite side of the room. In the meantime it has cooled somewhat and begins to sink toward the floor. Eventually it will flow into the area of the radiator again to be reheated, and the process repeated. The net effect of this is that convection currents are set up, as shown in the accompanying drawing, in the room; the practical effect is that heat is distributed over the room, and the air is warmed fairly uniformly. The general path of the air currents is upward at the source of heat and downward at points remote from it. One should note that it is the heated matter and not simply the heat itself that rises. Were it not for convection, the problem of warming a room would be rather difficult, since the air itself is such a poor conductor of heat.

Air convection on a large scale takes place in the earth's atmosphere. Great quantities of heated air rise into the upper atmosphere in the region around the equator, where the air expands and cools and finally settles back to the earth at some distance on either side of the equator. Eventually some of it flows into the equatorial region, again to repeat the cycle. This motion is modified, of course, by such factors as the rotation of the earth, the uneven topography of the land, and local warm

and cold spots. The results are all the complex motions of the atmosphere that determine our weather and climate.

Everyday examples of convection within liquids are not quite so common as they are in the case of gases; however, the principle is just the same. The hot-water heating system that is sometimes used for house heating is a good illustration. In this type of heating system the furnace boiler and radiators are filled completely with water. As the water is heated in the boiler, it rises by convection and flows through a system of pipes into the radiators where it gives up some of its heat and becomes cooler. The cooler and heavier water then returns to the boiler through another system of pipes. In this manner heat is transferred from the boiler to the radiators by a continuous convection current of water.

Convection currents are sometimes set up in large bodies of water by local heating in one region. The warm water expands and flows away from the heated area. Cooler water flows in to replace it, and the result is a continuous flow within the body of water. Some of the well-known currents in the oceans, such as the Gulf Stream, are partly a result of convection.

Expansion and Contraction

We have observed that an increase in the temperature of a volume of gas produces an increase in the internal pressure within the gas, and the result is a tendency for the gas to expand. In fact unless the gas is confined in a rigid container, it will expand, as was illustrated in the case of the heated balloon. The molecular explanation of this effect is fairly simple, since the increase in pressure is merely the result of increased bombardment of the molecules. On the other hand, if the temperature of the gas is lowered, the opposite effect will take place, and the gas will tend to contract.

Liquids have a tendency to expand and contract in much the same way as gases. Since liquids are generally in some kind of rigid container, the expansion takes place at the upper surface only, so that the level of the liquid rises as its temperature is increased and falls with a decrease in temperature. A typical example of this effect is the ordinary thermometer. However,

should a liquid completely fill a sealed container and be heated, expansion must take place. This will occur even at the expense of cracking the container, since a liquid is not nearly so compressible as a gas, and enormous pressures would be built up by the heat applied. In all hot-water-heater systems (not steam), provision for this expansion must be made; otherwise the boiler, radiators, or pipes would crack each time the water was heated. The expansion is conveniently taken care of by connecting a pipe to the system and letting it extend into an open expansion tank, located somewhere above the upper floor to be heated. The additional volume of water produced by the expansion flows into the tank, and very little additional pressure is created.

As a general rule, solids expand with increase in temperature and contract when they are cooled. They do not behave quite so uniformly in this respect as liquids and gases do, particularly for wide variations in temperature. For most practical purposes it is usually necessary to know how only one dimension of a solid will vary with temperature changes. This one-dimensional variation is called the linear expansion of the solid and may be expressed as the percentage change in its length for one-degree change in temperature. The following table gives this value for some common materials.

Material	Per Cent of Change in Length per °C.
Aluminum	0.0023
Brass	0.0019
Copper	0.0017
Cast iron	0.0011
Steel	0.0010
Ordinary glass	0.0009
Pyrex glass	0.0003
Fused quartz	0.00005

The problem of expansion is a very important one in structural engineering. The change of temperature with the seasons, or even with day and night, causes such materials as railroad rails, concrete pavements, and skyscraper beams to expand and contract considerably. As a result "expansion joints" must be provided in any large structure. These joints consist of some kind of connection or coupling which provides a gap that will

accommodate an increase in length. Without such gaps, the enormous force of expansion would cause structures to bend and buckle.

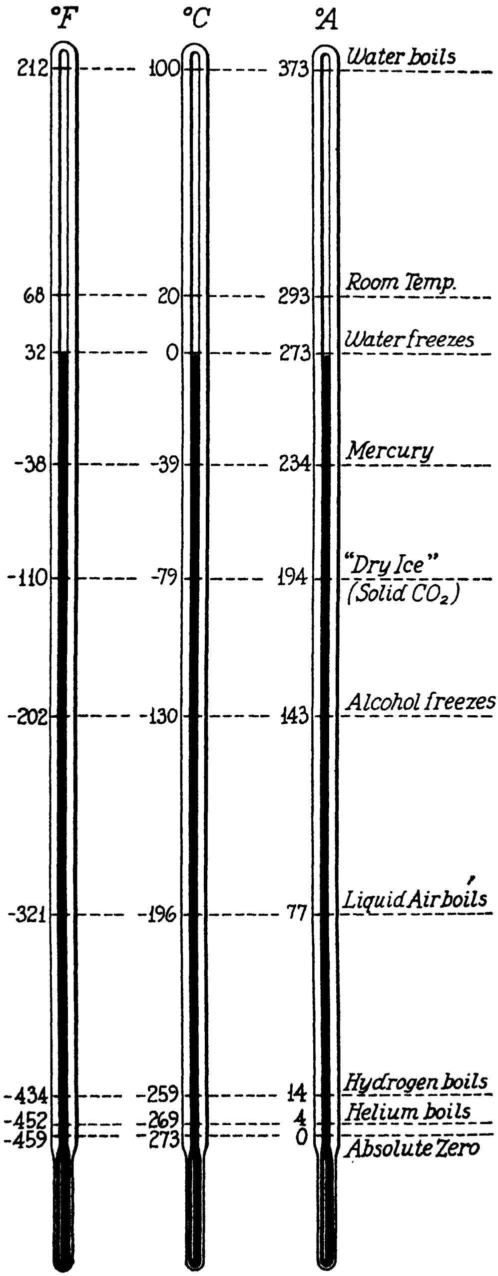
As a practical example of the need for expansion joints, let us calculate the change in length of one mile of railroad track from winter to summer. In some climates the temperature change may be as great as 60°C. From the table above, we see that steel will change in length by a factor of 0.001 per cent for each degree centigrade. For a change of 60°C. this will be 0.06 per cent, and 0.06 per cent of one mile is 3.16 feet, or about 38 inches. It is evident, therefore, that the tracks would be ruined if provision were not made for taking care of this expansion.

Many people have learned from experience that pouring boiling water into an ordinary drinking glass is almost certain to cause the glass to break, particularly if its walls are thick. The sudden increase in temperature on the inside of the glass causes the inside surface to expand more than the outer surface, thus setting up a stress in the glass that is great enough to break it. Pyrex will stand sudden temperature changes much better than ordinary glass, because its expansion rate is only one-third as great as that of glass. Fused quartz has such a low expansion rate that it will withstand almost any kind of a temperature change without breaking.

Measuring Temperature

We have just seen that the expansion of a substance can be calculated when we know how much its temperature is going to increase. The reverse of this process is also possible; *viz.*, when the amount of expansion for a given substance is known, its temperature increase can be calculated. This is the principle employed in the thermometer. We measure the expansion of a liquid in a tube with a scale graduated in degrees of temperature change and based upon the amount that the liquid expands with an increase of one degree in temperature.

The operation of the ordinary thermometer is quite familiar to most people. As is shown in the illustration, a cylindrical reservoir called the "bulb" is attached to the end of a glass tube having a small internal diameter. The bulb and a part of the tube are filled with a suitable liquid. (A number of different



Comparison of Fahrenheit, centigrade, and absolute thermometer scales.

Calibrating thermometers at low temperatures at the National Bureau of Standards. (Science Service photograph.)

kinds of liquids are used in household thermometers, but for laboratory thermometers mercury is generally employed.) Any expansion or contraction of the liquid in the bulb causes the top of the column in the tube to move up or down. The height of the column, therefore, is an indication of the relative temperature of the bulb.

In order to use the thermometer for temperature measurements, it is necessary to measure the height of the mercury column in terms of a standard temperature scale. Two such scales are in general use at present. One is the "Fahrenheit" scale which expresses temperature readings in "degrees Fahrenheit"

and is usually abbreviated °F. It is commonly used for measuring temperatures associated with the household, the weather, and industrial work. The other is the centigrade scale which expresses temperatures in "degrees centigrade," and it is abbreviated °C. This scale is almost universally used for all kinds of laboratory work. Both scales are arbitrary, of course, but the centigrade is much more convenient. The only reason why the Fahrenheit is still in use is that it was well established when the centigrade scale was invented.

In the case of each one of these temperature scales, two arbitrary temperatures were selected that were sufficiently far apart to permit of several divisions of degrees of temperature between them. In establishing the Fahrenheit scale, a rather long and complicated series of steps was worked out whereby the temperature of freezing water was set at 32°, the temperature of the human body at 98.6°, and the temperature of boiling water at 212°. The centigrade scale was established by taking the freezing and boiling temperatures of water at sea-level pressures as the two arbitrary temperatures. These are temperatures that may be readily obtained by anyone wishing to verify the readings. The temperature of freezing water was designated zero; that of boiling water was called 100°. Each of the 100 equal divisions on this scale represents 1°C.

The comparative relationship between the temperature readings on the Fahrenheit and centigrade thermometers may be observed from the drawing just referred to. Note that the difference between the freezing and boiling points of water on the Fahrenheit scale is 180 Fahrenheit degrees and that on the centigrade scale it is 100 centigrade degrees. It is seen, therefore, that the centigrade degree is equal to 1.8 Fahrenheit degrees, and this constant is always used in transferring centigrade readings into Fahrenheit readings, or the reverse. An additional calculation must be made, however, to get the two zero points equated. Since the freezing point of water is 32° on the Fahrenheit scale and zero on the centigrade scale, 32 must be added to the figure after converting from the centigrade reading to Fahrenheit, and 32 must be subtracted from the Fahrenheit reading before converting to centigrade. The normal temperature of the human body is 37°C., and it might prove of

interest to the reader to see if he can proceed through the relatively simple mental processes of calculating this figure from the Fahrenheit temperature of the body stated above.

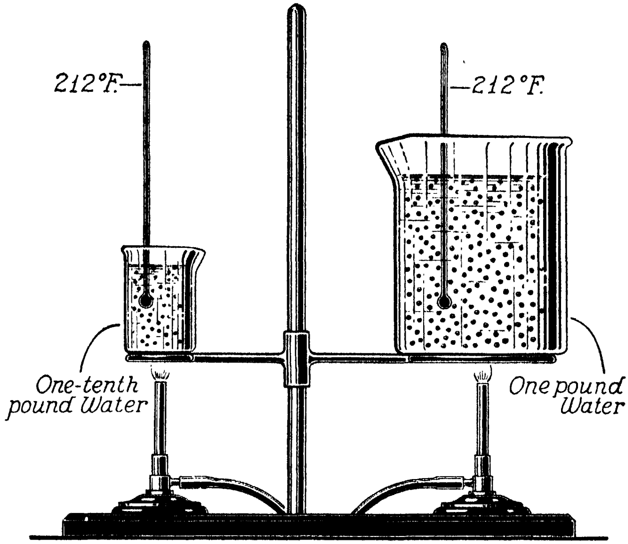
In the case of each of these thermometers the zero point represents a definite temperature, and the thermometer reading is simply a comparison of the temperature measured with the temperature corresponding to the zero. When a temperature lower than zero is to be measured, it must be expressed as a negative number.

Earlier in this chapter we pointed out that there is a limit to how low temperatures can go. This limit is theoretically 273.1 degrees below zero on the centigrade scale (−273.1°C.), or the absolute zero. It represents the condition where molecular motion would cease entirely. For the purpose of making calculations in problems dealing with molecular motions, it is often convenient to express temperatures with reference to absolute zero. This involves using a temperature scale the zero of which corresponds to −273.1°C. Such a scale can be obtained by adding 273.1 to centigrade thermometer readings. It is illustrated in the drawing referred to above and is called the absolute, or Kelvin, scale. Bear in mind, of course, that no such thermometer exists. It would be physically impossible to build an ordinary thermometer having this range, because any liquid suitable for ordinary temperatures would freeze long before −273°C. was reached. When it is necessary to measure very low temperatures, electrical methods are generally employed.

Measuring Heat

In this enlightened age probably few people believe that "a watched kettle never boils," but we all know that it seems to take a longer time to heat water if one is in a hurry. We have learned by experience that one way to help the matter along is to heat just as little water as is necessary. A kettle half full can be brought to the boiling point in half the time required for a full one. Why should this be true when the temperature change is the same; *i.e.*, both are started at the temperature of the water from the tap, and both end up at 100°C.

This is no freak of nature, and the answer is quite simple when explained in terms of moving molecules. In both cases

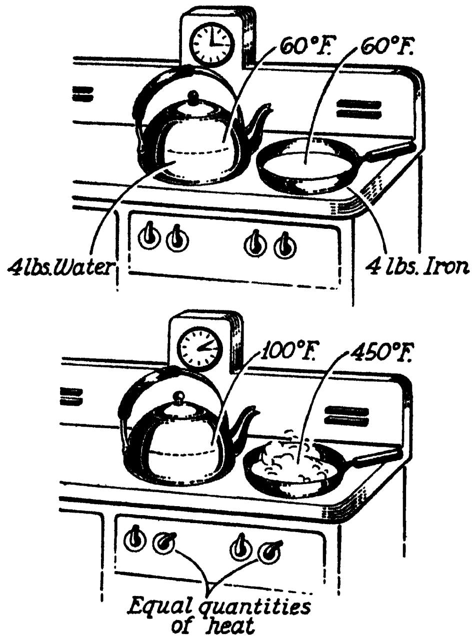


Although the temperatures of the two beakers of water are the same, the large one contains ten times as much heat.

the molecules are speeded up by the same amount, but in one case there are twice as many molecules to be speeded up; so it takes the same flame twice as long to supply the necessary heat energy. This all seems simple enough, but it illustrates the important fact that the temperature of an object does not tell us how much heat it contains. We must know also the weight of the object. Stating the same fact in another way, the temperature of an object tells us the average speed of its molecules, but we must weigh the object to know how many molecules are in the race. Ten pounds of boiling water contains ten times as much heat as one pound of boiling water, although both are at the same temperature. This fact is illustrated by the experiment shown in the accompanying drawing. The average speed of the water molecules is the same for each beaker because the temperature is the same; the larger beaker, however, contains ten times as much heat because it contains ten times as many molecules.

Thus we see that both the temperature and the weight of a body are factors in determining how much heat it contains. The third and final factor that must be considered is the kind of matter involved. For example, a pound of water and a pound of iron at the same temperature do not contain the same amount

of heat. Stated in another way, if equal quantities of water and iron contain the same amount of heat, the temperature of the two substances will be different. This is a fact of everyday observation in the household, as illustrated in the next drawing. If we consider an iron frying pan as weighing four pounds and the teakettle as containing two quarts of water, the quantities of iron and water are approximately equal. Now, when each is placed over an equally hot gas flame, the frying pan will become very hot long before the water has even started to boil.



The temperature of iron increases more rapidly because its heat-holding capacity is relatively small.

Here we have an example of equal quantities of heat added to equal weights of matter. The resultant temperatures, however, are quite different, because different kinds of matter are involved. In general, the amount of heat energy required to produce a given temperature change in a pound of matter will vary considerably, depending on the kind of matter. Apparently it is easier to change the average molecular speed in some substances than in others. This property of resistance of a substance to heating or cooling is technically known as its specific heat. For practical purposes this means its "heat-holding" capacity. Let us return again to the illustration of the frying pan and kettle of water and this time heat them both to the same temperature, say 212°F. Upon being allowed to cool, the pan will cool to room temperature long before the water. This is so because water has a higher heat-holding capacity than iron.

The heat-holding capacity, or simply heat capacity, is different for all substances. Since this is true, it is convenient

to have a quantitative means of expressing this difference. A logical method is to call the heat capacity of water 100 per cent and then compare the heat capacity of other substances to it. The method works pretty well, since water has the highest heat capacity of any of the ordinary substances. For comparison the approximate heat capacities of equal weights of some common substances are listed below:

Substance	Per Cent of Heat Capacity of Water
Water	100
Air	24
Aluminum	22
Iron	12
Copper	9
Ice	50
Rock	20
Wood	42

Now we can begin to understand why the thermometer alone does not tell us how much heat an object contains. In addition to knowing the temperature we must know also the weight of the object and its heat capacity; then the heat in the object can be expressed in terms of the heat in an equal weight of water. To get around the awkward way of expressing the result, there has been adopted a unit of heat called the "calorie." It is defined as follows:

1 calorie = heat required to raise the temperature of 1 gram of water 1°C.

The use of the calorie simplifies greatly the measurement of heat. If a gram of water is heated from 20°C., which is about room temperature, up to about 100°C., or boiling temperature, 80 calories of heat will be required. One gram of iron heated by the same amount will require 12 per cent of eighty, or only 9.6 calories. When they both cool off, the water has to give up 80 calories, whereas the iron must part with only 9.6 calories. This explains why water seems to "hold" heat better than iron; it also shows exactly how heat is measured. In any future dealings with this object, it is important to remember that temperature and heat do not mean the same thing, as has been noted. One is measured in degrees, and the other in calories; one is a

measure of the average speeds of the molecules, and the other takes into account also the total number of molecules.

The calorie is the universal unit of heat used in scientific laboratories, all over the world. In the United States and Great Britain, however, it never has been used very extensively in commercial work. Instead, a unit of heat called the British thermal unit, abbreviated B.t.u., is employed. It is defined as follows:

1 B.t.u. = heat required to raise the temperature of 1 pound of water 1°F.

Anyone who has ever had occasion to read an industrial report on heating, fuels, or refrigeration, written in this country or in England, has probably found that the reference of heat measurement was in B.t.u.'s rather than in calories. Also the temperature measurements will be in Fahrenheit rather than in centigrade degrees; and the weight measurements will be in pounds, not in grams. This is merely a difference in the system of measurement and does not in any way modify the principles that we have discussed above. Heat quantities measured in either system may be compared by using the following relationship:

1 B.t.u. = 252 calories.

The table of heat capacities given above applies to either system of units and may be used with equal exactness and facility whether the unit employed is the calorie or the B.t.u.

To some readers, all the preceding discussion might have a little more meaning if it were illustrated with one practical example. This is in order, particularly since hundreds of similar examples must be solved every day for our comfort. Suppose that a room having a volume of 2,000 cubic feet is to be air conditioned. The temperature is to be kept at 70°F. when the outside temperature is 100°F. How much heat must be extracted from the incoming air in order to maintain this lower temperature? Ordinary air weighs about 0.075 pound per cubic foot. The weight of the air in the room, then, is 150 pounds. The change of temperature required is 30°F. The heat capacity of air, as was noted in the table above, is 24 per cent. Therefore the total heat that must come out of the room full of air is equal to:

$$\underset{\text{(weight)}}{150} \times \underset{\text{(change in temperature)}}{30} \times \underset{\text{(heat capacity)}}{24 \text{ per cent}} = 1080 \text{ B.t.u}$$

Thus the air-conditioning engineer is given something definite on which to base his design. If he decides that this air must be changed four times per minute, he must provide a refrigeration system that can extract four times this amount, or 4320 B.t.u., per minute from the incoming air. Since B.t.u.'s of refrigeration can be figured in terms of dollars and cents, we have a basis for both the cost of the installation and the cost of operating it.

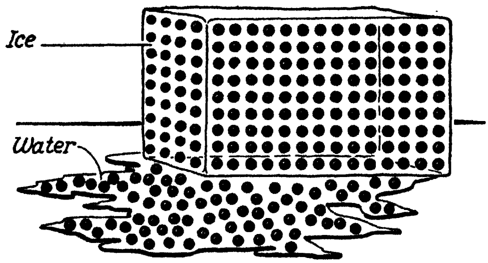
The Melting Process

Everyone has observed that many solids melt and change into the liquid state when their temperature exceeds a certain value. The exact temperature at which melting takes place will depend upon the kind of substance, but there is a definite temperature for all substances at which the melting process starts. This is known as the melting point of the substance. One of the most common examples of this fact is ordinary ice which has a melting point of 32°F., or 0°C.

Although melting is accomplished by adding heat to a substance, it is an interesting fact that the temperature of the substance does not increase at all after the melting point has been reached. For example, if a vessel filled with cracked ice is set on a hot radiator and allowed to melt, the temperature of the mixture of ice and water will remain at 32°F. until all the ice has been melted. Apparently the heat that is absorbed by the mixture from the radiator is used for some purpose other than to raise its temperature. The explanation for this fact can be found from a study of the changes of molecular conditions that take place within a substance during the melting process.

The accompanying drawing shows a "molecular view" of a cake of melting ice with the molecules enormously magnified. It is seen that the molecules in the solid cake have a uniform arrangement in a definite pattern, not greatly unlike the pattern of black squares on a checkerboard. The arrangement does not change appreciably as long as a substance remains in the solid state. This does not mean that the molecules cannot move; on the other hand, it does mean that their motion is limited to a

kind of vibration around one central point. This type of molecular arrangement is what gives solids definite boundary and form.



Molecules in liquid state have greater freedom of movement.

Now compare the molecular conditions found in the ice with those in the pool of water around it. Here the molecules do not have any definite arrangement with respect to one another, and each molecule is free to drift about anywhere within the liquid. This does not mean that the average speed of the molecules in the water has been increased; rather it means that in the liquid state they have greater freedom of movement.

The molecular arrangement shown in these diagrams, of course, is based upon inference from experiments rather than direct observation, for there is no known way of making molecules visible. However, experiments do indicate that the essential difference between the solid and liquid states is one of molecular arrangement and degree of freedom. In the solid state, each molecule is more or less confined to one small region, whereas in the liquid state each is free to migrate to any point within the entire volume of liquid. Simple experiments show that the change from a solid to a liquid absorbs heat without increasing the temperature of the substance. Therefore melting must be a process of breaking down the forces that hold molecules together in a solid, rather than a process of increasing molecular speed.

The heat energy used in the melting process is called the latent heat of fusion, wherein the term "latent" implies that the heat cannot be recovered from the liquid until it returns to the solid state. Latent heat for most substances has been measured accurately. The latent heat for ordinary ice is approximately 144 B.t.u. per pound, which means, of course, that 144 B.t.u. of heat is required to melt one pound of ice. Latent heat provides for ice its best known and perhaps most important use in various kinds of refrigeration. It is important to note, however, that ice cools its surroundings not because it is cold but because it melts. The mere presence of a cake of ice in a refrigerator will

not produce any appreciable cooling. It is the melting of the ice that cools the refrigerator, since the surrounding materials (food, liquids, etc.) must give up 144 B.t.u. of heat for every pound of ice melted. By giving up this heat, the materials are soon cooled to the desired temperature, and in the meantime some of the ice has disappeared, of course, and must be replaced as needed.

We have discussed the basic facts about the melting process but have said nothing about the closely related process of freezing. This is the term applied to the change from the liquid back to the solid state. As a matter of fact, the only difference between these two processes is the direction of flow of latent heat. For example, when one pound of ice melts, 144 B.t.u. of heat flows into the ice in order to produce the change of state; and when one pound of water freezes, 144 B.t.u. of heat flows out of the water as it solidifies. Aside from this important distinction, an explanation of one process should provide a clear understanding of the other.

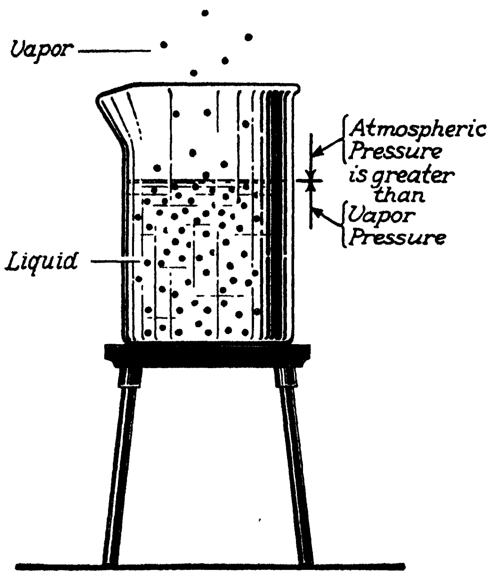
Our explanation of the melting process has been illustrated with water as a typical substance. The same type of explanation will apply to most other substances, except, of course, for differences in melting points and latent heat values.

Evaporation

When water is exposed to the atmosphere in an open vessel, it gradually disappears by the process we call evaporation. It is probably equally well known that this process is always accelerated by an increase in the temperature of the water. The same conditions hold true for other liquids, although the rate of evaporation may vary considerably for each. Alcohol and gasoline, for example, evaporate faster than water, whereas mercury does not evaporate nearly so rapidly. Apparently, therefore, the rate of evaporation depends upon both the temperature and the kind of substance being evaporated. This is true because both are factors in determining the upward pressure at the surface of a liquid.

In the discussion on melting, it was noted that the molecules in a liquid have considerable freedom of movement. However, a liquid does have a definite boundary at its surface; this is true partly because there is an attraction between the molecules

that tends to hold them together and partly because the combined forces of gravity and atmospheric pressure tend to keep the surface flat. Because the molecules of the liquid are in rapid motion, there is a tendency for them to break through the surface, in spite of the downward pressure of the atmosphere. Occasionally a molecule of the liquid strikes the surface with enough energy so that it breaks through and escapes into the atmosphere above, as illustrated in the drawing. This process of gradual molecular escape is what constitutes evaporation.



Evaporation involves slow escape of a few energetic molecules.

The tendency of molecules to escape from the liquid manifests itself as a steady upward pressure known as the "vapor pressure" of the liquid. For any given substance the vapor pressure increases with temperature, a condition that would be expected, since an increase in temperature increases the speed of the molecules of the liquid. As an illustration, at 50°F. the vapor pressure of water is about one per cent of atmospheric pressure, whereas at 200°F. it is almost 80 per cent. Similarly, the vapor pressure for different substances varies widely. For example, the vapor pressure for water at 86°F. is 4.4 per cent of atmospheric pressure; for alcohol at the same temperature it is 10.4 per cent; and for ether it is 85 per cent of atmospheric pressure. This means that alcohol will evaporate considerably faster than water, and ether will vaporize almost as fast as it is poured from the container.

As a liquid evaporates, the remaining portion, or its surroundings, gradually becomes cooler. The effect is quite noticeable when a rapidly evaporating liquid such as ether is poured

on one's hand. It is also equally noticeable when an open container of a rapidly evaporating liquid is insulated enough so that heat from the outside does not warm it up rapidly. The explanation for this cooling effect of evaporation lies in the fact that molecules must have an increased amount of energy in order to escape from the liquid during evaporation. When this increased energy is absorbed from the surroundings, the latter are thereby cooled, as happens in the case of ether evaporating from the hand. In case the liquid is in an insulated container, only the most energetic molecules escape from the liquid during evaporation; therefore the average speed of those remaining is gradually reduced, so that the temperature of the liquid is lowered.

The cooling effect of evaporation serves as a temperature regulator for many of nature's processes, as well as for some that have been devised by man. One of the most noted natural examples of this regulation is the cooling effect of the evaporation of perspiration from the human body. Excessive heating of the body in warm weather causes perspiration, and the evaporation of this perspiration absorbs heat from the body and produces the necessary cooling to help restore the body temperature to normal. Artificial applications of this principle include uses in refrigeration plants, air-conditioning systems, and similar cases where the evaporation of a liquid is involved and where its cooling effects can be utilized.

What happens to the vapor that escapes from an evaporating liquid? Does it remain in the atmosphere indefinitely? The answer to these questions depends upon the degree of saturation of the atmosphere with the molecules of the liquid. Water vapor is always present in variable amounts in the spaces not occupied by the other constituents of the atmosphere. One should bear in mind, however, that water is not an essential part of the atmosphere and that it does not enter into any chemical combination with the molecules of the other gases. We measure the water-vapor content of the atmosphere in terms of the percentage of complete saturation for a given temperature. This percentage figure is called the "relative humidity" of the atmosphere. To illustrate its meaning, if the relative humidity is 50 per cent, the space in the atmosphere contains only one-half as much water vapor as it could hold at that temperature. If, however,

the relative humidity is 80 per cent, this same space is four-fifths saturated for that temperature. Bear in mind the fact that the temperature must be specified in each case, a condition necessary because the amount of water vapor that can be stored in the atmosphere increases with an increase in temperature. Therefore, as the temperature increases, the relative humidity is lowered, although the actual amount of water in the atmosphere remains the same.

What happens when the relative humidity reaches 100 per cent? This means, of course, that the atmosphere is saturated with water vapor, and it would appear that under this condition evaporation would be impossible. From a practical standpoint this is true; however, what actually happens is that any additional evaporation is completely counterbalanced by an equal flow of molecules from the atmosphere back into the liquid. Therefore, the net amount of molecules added to the atmosphere is zero.

Under ordinary circumstances a condition of 100 per cent relative humidity is brought about most frequently by the temperature of a portion of the air being lowered through contact with cooler surroundings. The temperature of the water-vapor molecules occupying the space within the air is lowered, of course, at the same time. The most common example of this effect is the formation of water droplets on the outside of a cold-water pipe or on a pitcher of ice water. The temperature of a film of air adjacent to the cool object is lowered to a point where some of the water vapor must return to the liquid state. The temperature at which this process begins is called the dew point for the atmosphere, and its value depends, of course, upon the relative humidity.

An example of condensation on a big scale is the formation of clouds. Warm air near the surface of the earth acquires large amounts of water vapor as a result of evaporation from the ocean and smaller bodies of water. As this air rises to the upper layers of the atmosphere, it comes into contact with much cooler air and eventually may be cooled below the dew point. At that time some of the water vapor is condensed out in the form of fine droplets. The tiny particles floating together in the atmosphere form clouds, and under proper conditions these

particles collect in larger groups as raindrops and fall to the earth in the form of rain. Should the condensation take place at the earth's surface, the result is the formation of a cloudlike blanket called fog.

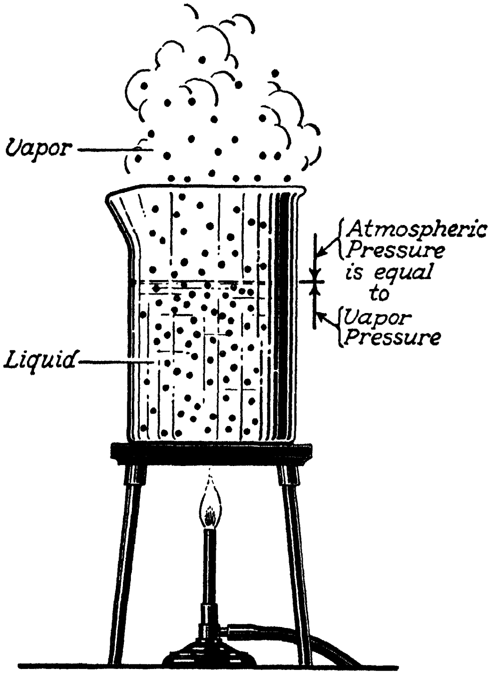
It is seen, therefore, that the processes of evaporation and condensation are responsible for maintaining the exchange of water between ocean and land. The heat energy for producing the evaporation of water is supplied by the sun, and convection currents within the atmosphere are responsible for providing the necessary circulation of moisture-laden air. The heat energy is liberated to the cooler upper atmosphere when condensation takes place, and the water then falls to the ground. This cycle is essential to the growth of land plants and to the existence of life on land.

Boiling

We have noted that the rate of evaporation increases rapidly with an increase in temperature. This condition is brought about by an increase in the vapor pressure of the liquid as the temperature rises. When a liquid is heated sufficiently, its vapor pressure will increase to the point where it is equal to the pressure of the atmosphere, and the temperature at which this condition occurs is called the boiling point of the liquid. The boiling point of water being 212°F., for example, means that at that temperature the upward pressure of water molecules is great enough to counterbalance the restraining pressure of the atmosphere. Accordingly, all the molecules are free to escape from the liquid without hindrance, and they proceed to do this just as rapidly as heat is supplied to the boiling liquid. An attempt has been made to visualize in the drawing on the following page the molecular activity in a beaker of boiling water.

The boiling process is similar to evaporation in that it involves an escape of molecules from the liquid, but it differs from evaporation in that it takes place much more rapidly. In evaporation the change from liquid to vapor takes place at the surface only, whereas in the boiling process the vapor may form anywhere within the liquid. This accounts for the formation of bubbles within the interior of a vessel of boiling water, although no such bubbles form during evaporation.

The boiling process has some features in common with the process of melting. The temperature of a boiling liquid remains constant until the entire volume of liquid has "boiled away," or changed to the vapor state. In the case of boiling water, the temperature will not exceed 212°F., regardless of how rapidly heat is applied to it. The additional heat energy added to the liquid is utilized in breaking down at a more rapid rate the forces that hold the molecules in the liquid state and in giving them the freedom of molecules of a gas. Their average speed is not increased during the process; it is merely that more molecules within a unit of time have been given this speed. The water does not get hotter with added heating; it only boils away faster.

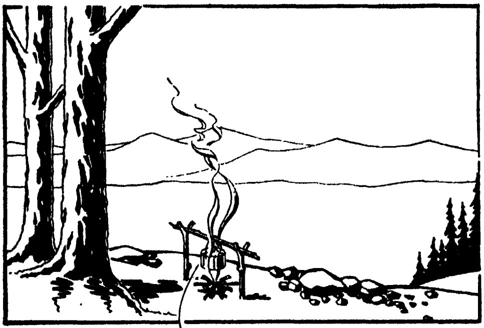


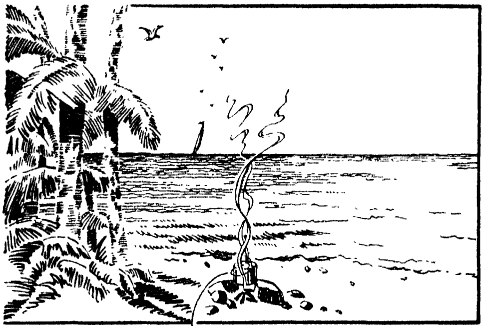
Boiling involves rapid escape of all molecules.

Another similarity between boiling and melting is the fact that they both involve definite quantities of heat. It has already been noted that 144 B.t.u. is required to melt a pound of ice. About 970 B.t.u. is required to evaporate one pound of water at the boiling temperature. This figure is known as the latent heat of vaporization for water. Incidentally, the heat required to evaporate a pound of water at temperatures below the boiling point is approximately the same as this value, but it does vary somewhat, depending upon the temperature.

It is seen, therefore, that both the melting and the boiling processes take place at a constant temperature and both involve the absorption of heat. They represent the two important changes of state for any substance.

There is one important difference between melting and boiling, and this involves the effect of pressure on these two changes of state. Changes in pressure have a negligible effect





The boiling temperature of water is lower on a mountain top than at sea level.

upon the melting temperature of a solid, but the boiling temperature of a liquid is definitely determined by the pressure at its surface. This is true, of course, because any change in this pressure means that a corresponding change in the vapor pressure of the liquid must take place before boiling can occur. The boiling temperature of water is 212°F. for an atmospheric pressure of 14.7 pounds per square inch, which is approximately equal to the air pressure at sea level. The pressure of the atmosphere decreases at high altitudes, and on top of Mount Whitney at an elevation of about 14,500 feet, the atmospheric pressure is reduced to nearly one-half its value at sea level. Under this condition water boils at about 185°F. The temperature of "boiling water" may vary considerably, therefore, depending

upon the pressure under which boiling takes place, as is illustrated in the drawing. The boiling temperature of any other liquid varies with atmospheric pressure in the same general way that it does for water.

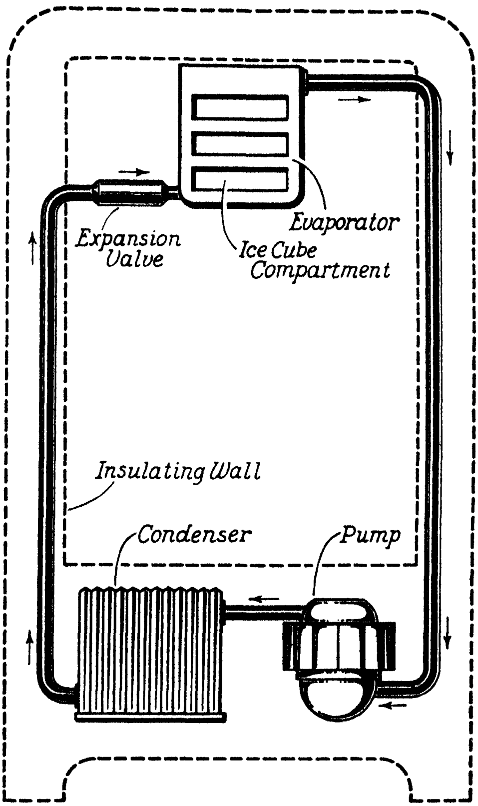
Water in the vapor state has many practical uses, a good example of which is the use of steam for heating houses, office buildings, and many kinds of enclosures that require artificial heating. The steam is generated in boilers at a centralized point in the system and is distributed to the radiators through a system of well-insulated pipes. While in the radiators, it is allowed to cool to the point where it condenses into the liquid state. As condensation takes place, the steam releases heat at the rate of 970 B.t.u. per pound. This passes through the iron walls of the radiator by conduction and furnishes heat to the room. The water in the radiator then returns to the boiler by a system of pipes and is converted into steam again. In this instance, the latent heat of steam provides a very useful means of transporting heat from one place to another.

The mechanical refrigerator common to many households is another piece of equipment that makes use of the transfer of heat by means of a vapor. In fact, there is no difference in principle between this type of refrigerator and a steam-heating system. The difference is only one of application. In the steam-heating system the latent heat of vaporization is given up to heat a room when steam condenses. In the mechanical refrigerator the latent heat of vaporization of a substance is absorbed from the refrigerator compartment to cool it as the substance evaporates.

Reference to the diagram of a typical household refrigerator may make clear how it operates. The refrigerator consists of four essential parts to provide the transfer of heat and thereby effect a cooling of the inside of the refrigerator, *viz.*, the pump, the condenser, the expansion valve, and the evaporator. The expansion valve and the evaporator are mounted inside the compartment of the refrigerator that is to be cooled. The pump and condenser, however, must be outside this compartment and exposed to the air of the room.

These four units are connected by a system of pipes which make a continuous loop, or closed circuit. The condenser is

partly filled with a special substance, usually referred to as the refrigerant, which must be a liquid with a low boiling point.



Heat absorbed by the refrigerant turning to a gas in the evaporator is removed from the insulated compartment, and given off to the outside air as the refrigerant turns back to a liquid in the condenser.

Sulphur dioxide is a suitable material for this purpose, since its boiling point is 14°F. at atmospheric pressure, a temperature that is 18° below the freezing point of water.

Let us see how the system operates to bring about a transfer of heat from inside the refrigerator to the outside. The pump, driven by an electric motor, generates a pressure in the condenser, and eventually this pressure becomes sufficient to condense the refrigerant into the liquid form. The liquid refrigerant is confined in the condenser and forced through the pipe to the expansion valve. The valve is adjusted so that it provides a

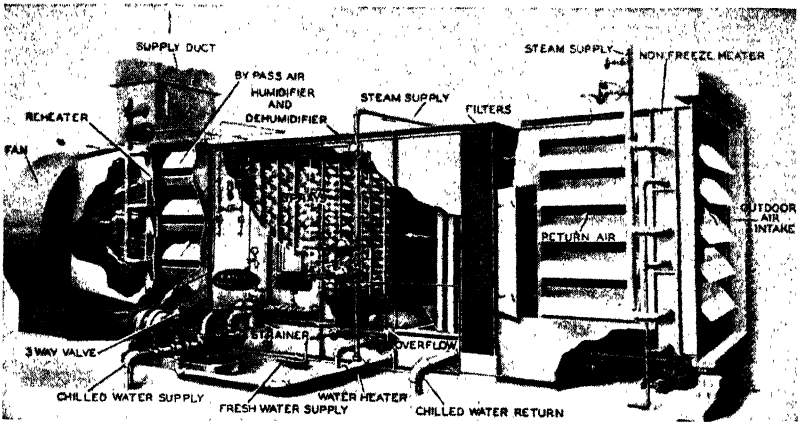
small needle-like opening between the condenser and the evaporator, thus permitting a fine spray of liquid to leak into the evaporator. The pressure within the evaporator is relatively low, and the liquid evaporates as soon as it enters this space. In changing state, however, the liquid must absorb heat. From its surroundings it takes up the latent heat of evaporation of the refrigerant, which is about 150 B.t.u. per pound for sulphur dioxide. The result is a cooling of the evaporator and the inside of the refrigerator. The vapor that collects in the evaporator is gradually drawn off by the pump, and the refrigerant is returned to the liquid state by the pressure in the condenser. In returning to the liquid state, it gives up its latent heat of evaporation to the surrounding air, incidentally heating the air somewhat.

It is seen, therefore, that the mechanical refrigerator is really a special kind of "heat pump." Some refrigerant is allowed to vaporize at the point to be cooled, absorbing heat from the surroundings; then the vapor is removed to another point outside the cooling system, and the heat is extracted from it by condensing the refrigerant under pressure. The process is continuous and may be carried on indefinitely.

Making Artificial Weather

In most climates people complain of the weather when it is hot and suffer from it when it is cold. Within recent years man has begun to do something about manufacturing indoor weather by conditioning the air to the point of greater comfort and more healthful living standards. The history of man's entire building program shows that he constructed buildings first for shelter, then for permanence, then for beauty, and finally for convenience. The next great advancement in building construction doubtless will be a wide-scale use of air conditioning to provide year-round comfort.

The modern science of air conditioning in its most elaborate form utilizes almost all the principles of heat so far noted in this study. Scientific research has established certain standards of maximum comfort for the atmosphere within a room. This involves optimum values of temperature and humidity as well as a minimum dust content. The function of a complete air-



Schematic drawing of a central station air-conditioning system in which air is drawn in from the outside as shown at the right, conditioned, then supplied to the required rooms through the duct shown at the left. (Courtesy of Carrier Corporation.)

conditioning system is to adjust all of these factors; it is much more than a means of providing "fresh air."

The details of operation of such a system are quite involved, but the important steps are as follows: (*a*) The air to be conditioned is collected from the outside atmosphere; in the larger systems a part of the same air is often recirculated, *i.e.*, it is used over and over again. (*b*) Filter systems are usually provided for removing dust, soot, and other solid particles. (*c*) Gaseous impurities such as carbon dioxide and sulphur dioxide are to a great extent removed by circulating the air through water. (*d*) Excess water vapor is removed by a process known as "dehumidifying." One method used is to cool the air to a point where the excess water condenses. (*e*) The temperature is adjusted to the desired point, depending upon the season. In the summer this involves cooling by a system of refrigeration, and in winter heat is added. In some cities the general standard for conditioned air is a temperature of about 80°F. and a relative humidity of about 50 per cent. The completely conditioned air is finally distributed through the rooms where it is to be used.

The design and operation of such a system involve careful attention to many principles of both mechanical and thermal engineering. The system must be designed to circulate a sufficient

volume of air for all conditions. The heat capacities and temperature changes of all materials involved must be carefully taken into account. The number of persons occupying the space to be supplied with air is an important factor, for the human body is a source of both water vapor and heat, and the average person may radiate heat at the rate of about 400 B.t.u. per hour; hence, in theaters and other places of public gathering the system must be designed to adjust itself to various sizes of audiences as well as to a wide range of outside temperatures. Contrary to popular supposition, air does not usually become stagnant because of lack of oxygen. The principal reason for a feeling of "stuffiness" is an excess of water vapor, a fact that emphasizes the necessity of dehumidifying the air in air conditioning.

REFERENCES FOR MORE EXTENDED READING

RICHARDSON, E. G.: "Physical Science in Modern Life," D. Van Nostrand Company, Inc., New York, 1939.

If you wonder how an airplane can fly, what limits the size and clarity of a television screen, and how many other commonly used devices work, you will find the answers here, in a book having the interesting style and thoroughness characteristic of so many British writers.

ERYING, CARL F.: "A Survey Course in Physics," Prentice-Hall, Inc., New York, 1936, Chaps. 5–7, inclusive.

For anyone who is beginning a study of the subject of heat these chapters provide an introduction that is interesting as well as informative. Fundamental principles are discussed and illustrated in a way that will not discourage the beginning student.

LEMON, HARVEY BRACE, "From Galileo to Cosmic Rays," University of Chicago Press, Chicago, 1934, Chaps. 14–19, inclusive.

The discussion of the subject of heat is quite complete, including brief historical sketches, fundamental principles, and applications from everyday life. Important principles are liberally illustrated with simple line sketches often with a humorous touch. This fact combined with the author's style of writing serves to hold the reader's interest even through the most difficult parts of the exposition.

BLACK, NEWTON HENRY: "An Introductory Course in College Physics," The Macmillan Company, New York, 1935, Chaps. XIV, XV, XVI.

This is a textbook designed for use in colleges. The treatment of the subject matter is complete, and the application of principles is well illustrated by numerous drawings of practical devices.

KNOWLTON, A. A.: "Physics for College Students," 2d ed., McGraw-Hill Book Company, Inc., New York, 1935, Chaps. XII, XIII, XVII, XXI.

In the language of the author this book is an outgrowth of an attempt to humanize and unify a course in general physics. It is characterized by a departure from the customary order of arrangement of subject matter in a textbook of physics. The material on heat provides a very good discussion for students who have had a little previous experience with the subject.

MILLIKAN, R. A., DUANE ROLLER, and E. C. WATSON: "Mechanics, Molecular Physics, Heat and Sound," Ginn and Company, Boston, 1937, Chaps. 9, 10, and 12.

This is an advanced text for the student who has had the equivalent of a good secondary-school course in physics. One outstanding feature is the inclusion of numerous plates and descriptive paragraphs on the early history of various developments in physics. Recommended for the serious student who is well past the beginner stage.

EWING, J. A.: "Thermodynamics for Engineers," Cambridge University Press, London, 1936.

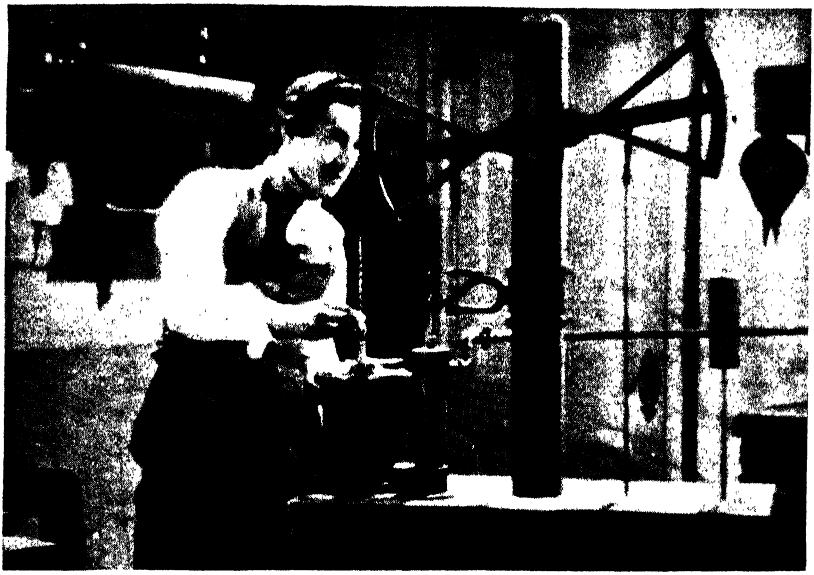
In spite of the advanced nature of the subject matter, much of the explanatory material can be followed by any student who has a general knowledge of heat and wishes additional information on thermodynamics.

The Advancement of Science, published by the British Association, Burlington House, London.

The British Association for the Advancement of Science reports its proceedings in this journal, containing science articles of general interest written in fine English style.

Heating, Piping and Air Conditioning, a monthly magazine published by Keeney Publishing Co., Chicago.

This is a technical magazine devoted to the design, installation, operation, and maintenance of heating, ventilating, and air-conditioning equipment in industry and in large buildings. It contains a section devoted to the activities of the American Society of Heating and Ventilating Engineers.



Johns-Manville.

9: LABOR SAVING

By the Development and Operation of Heat Engines

ONE characteristic of the human race is that people are always trying to get their work done for them, although few of us are willing to admit this fact. Nevertheless it is probably the chief reason why modern man has progressed further than his simian cousins. We do not know just when or where man made the first successful attempt to transfer some of his work to other forces. This information is buried with the fossils. There is little doubt, however, about the period of history during which one of the greatest achievements in reducing man's labor occurred. The story began in the year 1756; and it all started because a Scotchman was not allowed to run his own business.

The man in question was James Watt, born in Greenock, Scotland, in 1736. He was the son of a merchant who speculated

unwisely and lost most of his modest fortune when James was about nineteen years old. It became necessary for the young man to seek a means of support, so he made his way to London and became an apprentice to a maker of scientific instruments. The hard work and frugal living, necessitated in those days by this type of employment, made it impossible for him to complete his period of training. Instead, he returned to Scotland in 1756, planning to open an instrument-making business of his own at Glasgow. But the guilds in this profession failed to recognize Watt as a qualified instrument maker. He had not completed his apprenticeship, and they would not permit him to open a shop.

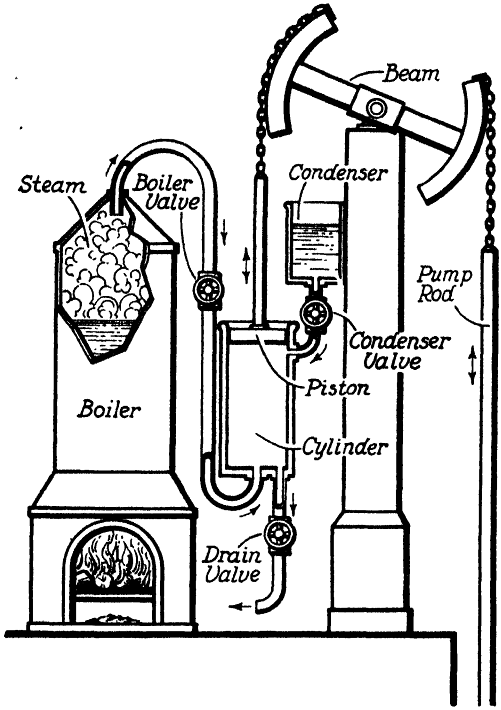
Just when Watt was at the point of giving up instrument making as a career, the University of Glasgow came to his rescue and gave him a job as custodian of laboratory apparatus. Among other duties he was assigned the task of repairing a working model of what was at that time the latest form of steam engine. In due time the repair job was completed. In working with the engine Watt was impressed with the fact that it was so crude and inefficient. He was sure that there must be a way in which it could be improved, and he set himself the task of finding out how it could be done.

Some of Watt's Contributions to the Steam Engine

In order to appreciate the magnitude of Watt's self-imposed task, we first should examine a few of the shortcomings of the steam engines of that time. They were all patterned after one that had been invented in 1705 by an English blacksmith, Thomas Newcomen. Their chief use was to operate pumps for pumping water out of mines.

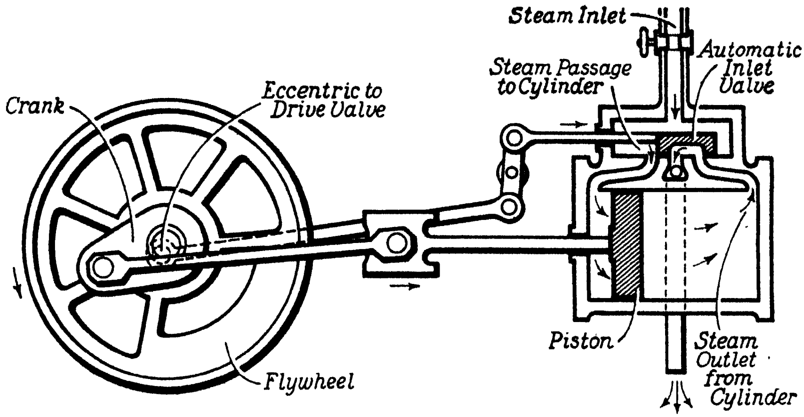
The general construction of the type of engine invented by Newcomen is shown in the accompanying drawing. A little analysis revealed that the four parts essential to its operation were a boiler, a cylinder, a piston, and a condenser. The piston was connected to the pumping mechanism by means of a horizontal beam pivoted in the center so that it could rock back and forth. A downward motion of the pump rod would pull the piston up to the top of the cylinder and a downward motion of the piston would lift the pump rod upward.

Operation of the engine would seem quite strange to a person accustomed only to the working of a modern steam engine.



Newcomen's engine.

The piston was normally held at the top of the cylinder by the weight of the pump rod. Steam was admitted into the cylinder by means of the boiler valve. When the cylinder was filled with steam, the boiler valve was closed, and the condenser valve opened. This procedure sprayed a jet of cold water on the hot steam inside the cylinder, causing the steam to condense rapidly. The condensation of steam within the cylinder produced a partial vacuum beneath the movable piston. It was thus forced to the bottom of the cylinder by the atmospheric pressure on its upper surface, thereby lifting the pump rod and operating the pump. After this process was completed, the valve at the bottom of the cylinder was opened so that the water produced by the condensation of steam drained out and new steam was admitted to the cylinder. The pressure of the new



Simplified drawing of modern steam engine.

steam equalized the atmospheric pressure, and the piston was returned to the top of the cylinder by the weight of the pump rod. This completed the cycle of operation.

This engine did not derive any useful work from the force of the steam itself. The useful work, *i.e.*, the lifting of the pump rod, was done by the atmospheric pressure on top of the piston. The function of the steam was merely to provide a gas that could be condensed rapidly so as to produce the required vacuum within the cylinder. For this reason these engines were sometimes called "condensing engines." They were very inefficient; and since all the valves were operated by hand, their cycle of operation was very slow. Yet they served a useful purpose in their day, and they mark the beginning of the application of steam to the production of mechanical energy.

It was a model of the Newcomen engine that Watt repaired in 1761, and he decided that the design could be improved. Thus he started on a career of invention that lasted for more than twenty years and resulted in several important changes in steam-engine design. We cannot discuss in this book all these inventions because they are too numerous and technical. We can, however, summarize the ultimate results of Watt's work by comparing Newcomen's engine with a drawing of a simplified modern steam engine.

From this drawing it will be observed that both ends of the cylinder are closed and that the connection to the piston is

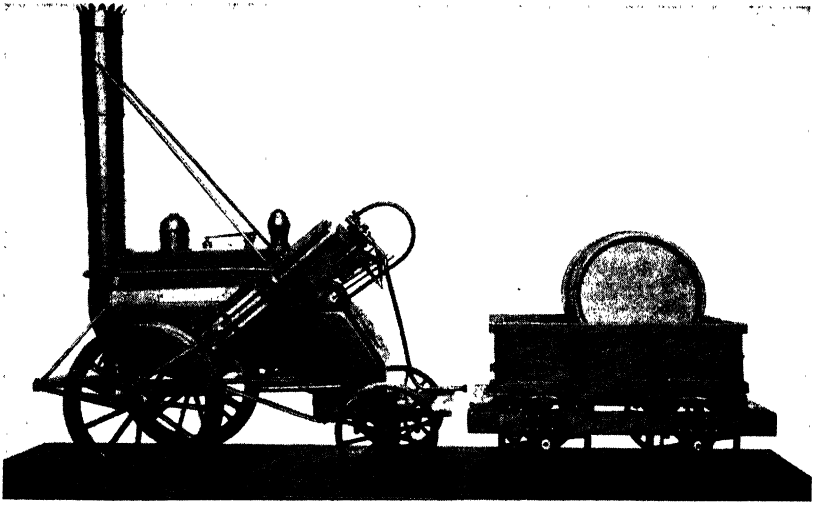
made by a rod that passes through a closely fitted orifice. Such an arrangement makes it possible to drive this rod back and forth with steam pressure applied first to one side of the piston and then to the other. Thus we have an engine operated entirely by steam instead of a combination of steam and atmospheric pressure. In order that this type of engine might operate smoothly and with any degree of speed, the admission of steam into either end of the cylinder had to be timed perfectly with the motion of the piston. Moreover, an outlet had to be provided for the used steam after it had expanded and driven the piston along the cylinder. This admission and release of steam in step with the motion of the piston was accomplished by means of a valve that was opened and closed by the movement of the piston itself. Thus the timing was always perfect, and the operation of the engine became entirely automatic. The automatic valve is often referred to as one of Watt's most important contributions to the steam engine.

Another fundamental improvement was the introduction of rotary motion. By means of a crank and flywheel, the back-and-forth motion of the piston was converted into rotating movement. This made it much easier to use the steam engine for driving all types of rotary machinery and extended its use to many fields other than the operation of water pumps.

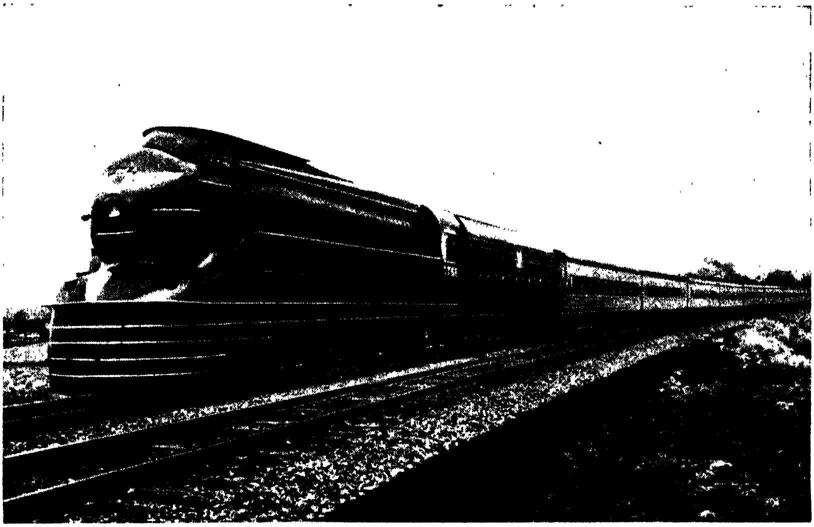
Universal Use of Watt's Engine

The improved form of the steam engine developed by Watt made it possible to replace the muscular energy of man and beast with the heat energy supplied by burning fuel. This led to a complete revision of methods of manufacture and transportation. The importance of Watt's work can best be judged by the fact that no fundamental changes in the design of the steam engine occurred for over a hundred years after his patents were granted.

Soon after its perfection the steam engine became a universal source of power for factories and mills. Such engines were called stationary engines, as they were permanently installed in one place. The installation consisted of a boiler for generating the steam and the engine for converting the heat energy of this steam into mechanical work. The rotary motion of the



The Rocket was the first successful railway locomotive ever built. (Science Service photograph.)



Modern streamlined steam locomotive is typified by this Pennsylvania engine. (Science Service photograph.)

crankshaft and the flywheel attached to it was transmitted by a system of belts from the flywheel to the machinery to be operated. This system constituted the standard method of obtaining factory power until the perfection of the electric motor a generation ago.

Early in the nineteenth century the steam engine was applied both to navigation and to transportation on land. Robert Fulton was not the first to consider propelling a boat by steam, as patents had been granted on this idea as early as 1736. He probably was the first inventor to apply steam power to practical navigation, and this he accomplished when he sailed the *Clermont* on her first voyage up the Hudson River from New York to Albany in 1807. Similarly, George Stephenson of England built the first successful railway locomotive, known as the *Rocket*, which made its maiden journey between Liverpool and Manchester in 1829. With the exception of a few Diesel-motored, streamlined trains and some electric locomotives, steam engines of Watt's general design still pull trains between cities and across continents.

Toward the latter part of the nineteenth century, when the electric dynamo was perfected, the steam engine was used to generate electricity. In a sense, therefore, James Watt contributed materially toward making possible the large-scale use of electrical energy. One of the tributes to his indirect accomplishments in this field is the fact that the internationally adopted unit for measuring electric power has been named the "watt."

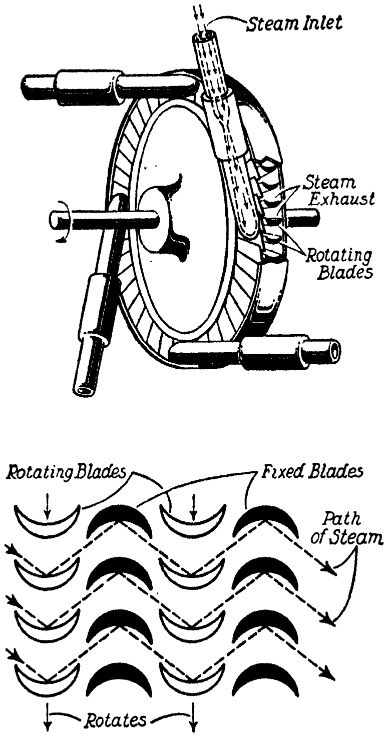
Steam Engines up to Date

The type of engine that has just been described is generally known as a reciprocating engine, because the piston moves back and forth. For some modern uses of steam power the reciprocating principle has definite disadvantages. In the first place there is an unavoidable loss of energy in starting and stopping the piston at the beginning and end of each stroke. Then, too, if high speeds are required, the rapid back-and-forth motion of the piston is a source of undesirable vibration; in fact reciprocating steam engines cannot for this reason run at extremely high speeds. Accordingly, for some types of steam-power application, they are gradually being replaced by a newer type, the turbine, which does not use a piston.

The turbine consists essentially of a paddle wheel driven by jets of steam coming from the boiler. In reciprocating engines, steam does its work by exerting an expansive force directly

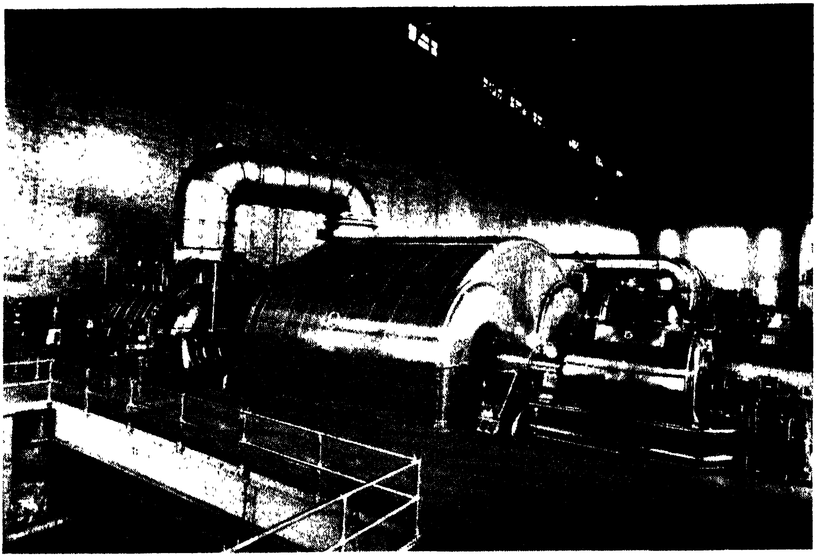
against the piston: whereas in a turbine, the escaping steam hurls itself against the paddles, or blades, of a rotating wheel. Since this jet of steam has a high velocity, it drives the blades forward with great speed by the force of the impact.

The principle of the turbine is illustrated in the top figure of the accompanying drawing. The rotating wheel, on which the blades are mounted, is called the rotor of the turbine. The tube from which the steam escapes is known as the nozzle. As the steam strikes the curved blades, a force is exerted on the rotor which starts it spinning around on its shaft. Thus the turbine principle makes it possible to convert the expansive forces of steam directly into rotary motion.



Single-stage steam turbine above, and arrangement of rotating and fixed blades for multi-stage turbine below.

The bottom figure of the drawing illustrates the construction of a so-called "multi-stage" turbine. The rotor consists of several sets of movable blades on a common shaft with each set separated by a set of fixed blades attached rigidly to the turbine housing. The fixed blades are curved in a direction opposite to the movable ones and their function is to guide the flow of steam as indicated. As the steam enters the left end of the turbine, it flows through the first set of movable blades and gives them a vigorous thrust in the downward direction; when leaving these blades, the steam is actually flowing upward. The first set of fixed blades, however, reverses the direction of the flow of steam so that another downward thrust is given to the second set of movable blades. The steam then strikes the second set of fixed blades, where it is again directed down-



A giant cross-compound steam turbine (9) used to drive an electric generator (5). (Courtesy of Consolidated Edison, New York.)

ward on the next set of movable blades, and the process is repeated to the end of the rotor. The diameter of each successive set of blades is made larger to accommodate the greater volume of steam as it expands. By the time the steam has reached the exhaust pipe at the right end of the turbine, it has been expanded to several times its original volume, and therefore most of the available energy has been extracted from it.

The original velocity of the steam escaping from the nozzles may be as great as 2,400 miles per hour. For this reason turbines sometimes develop the tremendous speed of 70,000 revolutions per minute. For most applications these high-speed units are connected to the load by means of reduction gears. Some types of slower speed turbines are coupled directly to high-speed electric generators, a condition that permits maximum mechanical efficiency.

The exhaust steam from a turbine is seldom discharged into the atmosphere. Instead, it is led into a device called a condenser, which consists of a vacuum chamber arranged to condense the exhaust steam and thus recover the heat that would be wasted if the exhaust were discharged into the open

air. The heat so recovered is used to warm the feed water going into the boiler. A condenser also eliminates the necessity of having available a large supply of fresh water for the boiler, as condensed steam can be used over and over again. This is very important in marine engines, since sea water would soon ruin the boilers.

The steam turbine in its practical form is a much more recent development than the reciprocating steam engine. The first practical turbines were introduced in England by Sir Charles Parsons about 1884. At that time, of course, practical reciprocating engines had been used for over a century. The first recorded account of the idea of a steam turbine, however, goes back to about 130 B.C., at which time Hero, a professor in the University of Alexandria in Egypt, invented a device later referred to as "Hero's reaction engine." In appearance and operation it was very similar to a rotary type of lawn sprinkler except that the rotation was produced by steam instead of water. However, the nature of the turbine principle is such that it can be made to work efficiently only when supplied with steam under high pressure. Since high-pressure boilers are a development of comparatively recent times, the development of the turbine has been delayed correspondingly.

One should not conclude from the foregoing discussion that the reciprocating steam engine is obsolete. There are some applications, *e.g.*, the railway locomotive, to which the turbine cannot be adapted easily; so the reciprocating engine will always be used to some extent. The increased efficiency and smoother operation of the turbine, however, have been responsible for its adoption in the electrical power plant field. Without the development of the turbine it would not have been practical to build large centralized plants for generating electricity except at natural water-power sites. This would have resulted in a much more limited use of electric light and power in both factory and home.

The Science of Energy Conversion

Our discussion of the steam engine has illustrated one general method of changing the heat produced by fuels into the mechanical energy of a moving machine. This is one example of what is

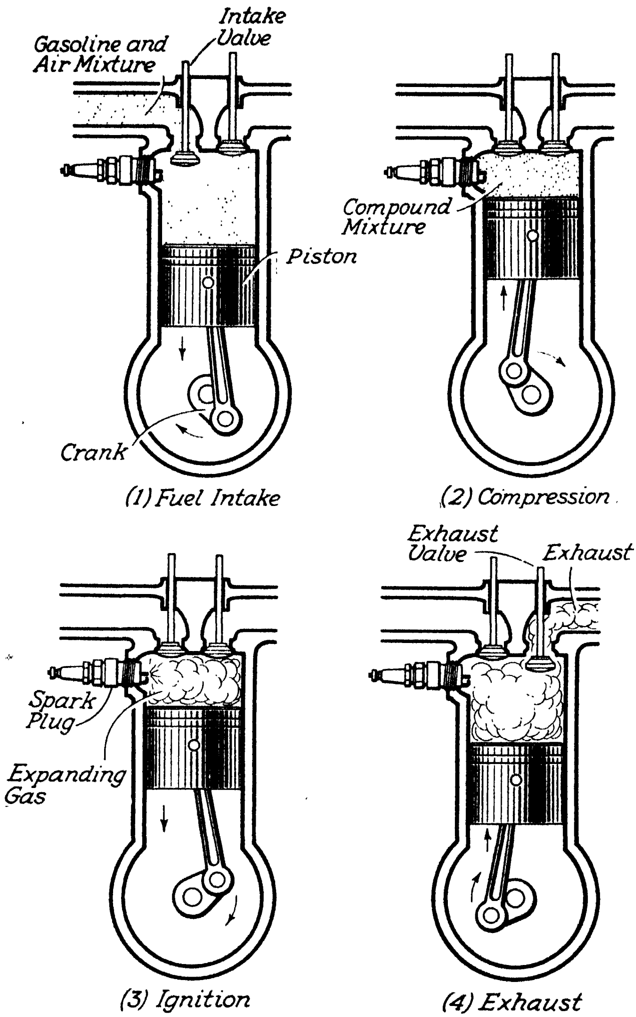
known in science as "energy conversion." The thing that we call energy exists on earth in a number of different forms, such as heat, light, electricity, and mechanical motion. Most scientific devices depend upon the principle that energy can be converted from one form into another. In the steam engine, for example, we have the following energy conversions taking place: First, chemical energy stored in fuel changes into heat energy of combustion; second, heat energy of combustion changes into energy of molecular motion in steam; and, third, energy of molecular motion changes into mechanical energy of moving piston or spinning rotor. Thus the energy really exists in at least four different forms during the process.

The essential principle of a heat engine lies in the fact that heated matter expands, and it is this expansion force which is harnessed and made to do useful work. In theory, therefore, any kind of substance that expands when heated is a potential heat engine. Steam happened to be one of the first substances that was used practically for this purpose because of its very early discovery and the ease with which it could be produced in quantity. Later on other substances were also used as we shall see in the discussion to follow.

Moving the Firebox into the Cylinder

So far we have discussed the type of heat engine that derives its energy from an expanding gas which has been heated outside the engine proper. There is another class, typified by the gasoline engine, in which the combustion takes place within the cylinder. This comprises the so-called "internal-combustion" engines. Historically they are a later development than the steam engine, and only within the last two or three decades have they offered important competition to steam as a source of power. Their most extensive application is in the automobile and airplane industries.

Firearms are perhaps the earliest example of the production of mechanical energy from internal combustion. Burning gunpowder produces a gas capable of exerting very quickly an expansive force, and this force drives the bullet from the gun barrel at high speed. A gasoline engine has some points in common with a gun or cannon, in that the piston is driven



Operation of the four-stroke cycle in a gasoline engine.

along the cylinder by the expansive force generated by burning a mixture of gasoline vapor and air inside the cylinder. The important difference, however, is that the motion of the piston must be harnessed, whereas that of the bullet is not; and the cylinder must be "reloaded" in a fraction of a second so that the operation of the engine will be continuous.

The method of accomplishing this sequence of operation is illustrated by the four diagrams of the gasoline engine. These

diagrams represent the four essential steps in the operation of one type of gasoline engine. Note that this is a reciprocating engine and that the arrangement of cylinder, piston, and crank are very similar to that of the steam engine; however, the driving force is applied to only one side of the piston. Intake and exhaust valves are provided for the admission of fuel and the removal of burnt gases. The fuel must be a substance that will burn exceedingly fast.

The details of each step illustrated by the diagrams may be outlined as follows:

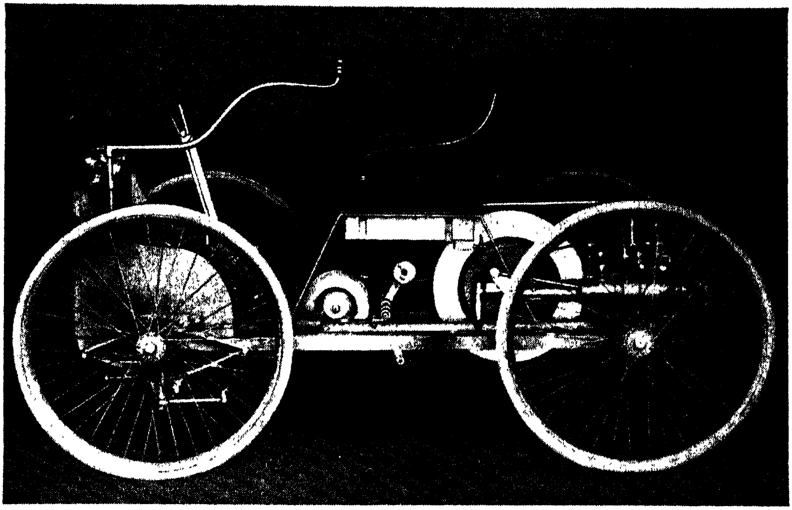
1. A mixture of gasoline vapor and air to provide oxygen is drawn into the cylinder through the intake valve by the downward motion of the piston. This valve closes at the bottom of the stroke.

2. The mixture is compressed to about one-sixth of its original volume by the upward motion of the piston, which produces a pressure of about 100 pounds per square inch.

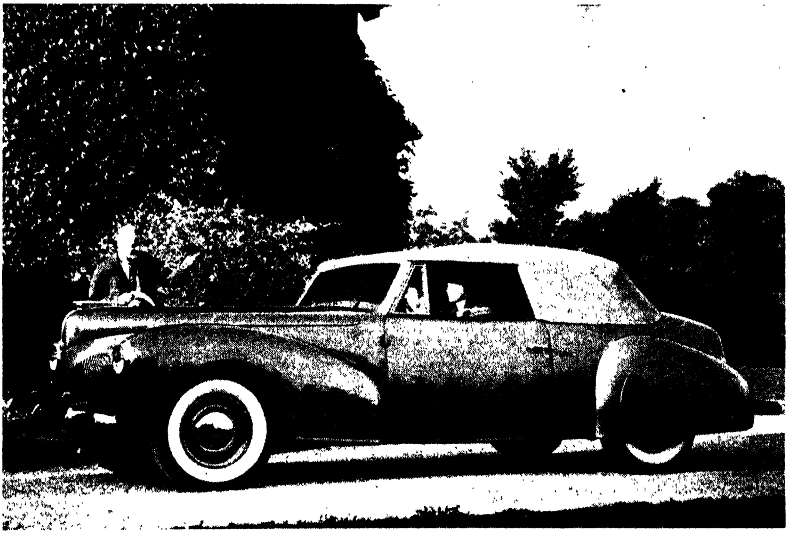
3. At the beginning of the second downward stroke, the compressed mixture is ignited by the spark plug. The burning of the gas builds the pressure up to about 400 pounds per square inch. The gases produced by the combustion expand with explosive force and drive the piston downward.

4. At the beginning of the second upward stroke, the exhaust valve opens. As the piston moves upward, the expanded gases are driven out through the exhaust pipe.

In this type of engine one power stroke is obtained for every two revolutions of the crankshaft, which means that four strokes of the piston are required for a full cycle of operation of each cylinder. In a single-cylinder gasoline engine this makes the operation rather jerky. To steady it, a large, heavy flywheel is mounted on the crankshaft. The inertia of the flywheel tends to keep the shaft rotating at a constant speed during the interval between power impulses from the piston. Modern automobile engines usually consist of four or more cylinders coupled to one crankshaft so that by properly spacing the individual cranks on this shaft, the energy impulses of each cylinder occur at different times. As a result, the flow of power is much more uniform, and a lighter flywheel can be used. This arrangement



The first Ford automobile built in 1904 had a one-cylinder gasoline engine with the flywheel and drive mechanism mounted under the seat. (Courtesy of Ford Motor Company.)



A modern automobile with an engine consisting of several cylinders is typified by this 1940 Lincoln Zephyr. (Courtesy of Ford Motor Company.)

also makes for smoother operation in a six- or eight-cylinder engine than is secured in a four-cylinder one.

One interesting fact about the gasoline engine is that all the successive operations of fueling, compression, firing, and exhaust must take place in a very short interval of time. In a typical modern automobile traveling at thirty-five miles per hour, each cylinder must go through its cycle about seventeen times per second. This allows only a few thousandths of a second for the opening and closing of valves, while the combustion and expansion must take place in one sixty-eighth of a second.

Perfect timing of these operations is necessary, therefore, for the functioning of a gasoline engine. The valves are driven by a camshaft geared directly to the crankshaft; and once they are properly set, they cannot easily get out of adjustment. The time of occurrence of the spark in each cylinder is controlled by a device called a "timer," or distributor, which is, in reality, a mechanically operated switch geared to the camshaft. The spark timing must be varied to suit the speed of the engine. So as to start the burning of gasoline at the proper instant, if the engine is running fast, the spark must occur earlier in the cycle than when it is running slowly. In the earlier models of gasoline engines used in automobiles a spark lever was provided for this purpose; it was manually operated so that the spark was advanced or retarded as needed. In modern automobiles, however, the spark control is made automatic by electrical and mechanical attachments.

Accessories to the Gasoline Engine

A gasoline engine cannot operate without certain accessory devices for the mixing of the fuel with air, the ignition of this mixture, and the removal of excess heat. Improvement of these devices has played an important part in making the modern automobile engine run smoothly and operate efficiently. Let us see, then, just how these accessory functions are performed.

As the gasoline comes from the supply tank, it flows into a device called a carburetor which is connected to the intake opening of each cylinder by means of a pipe line, or the intake manifold. As the intake valve opens, air is drawn through the carburetor. This air, rushing past a nozzle connected to the

gasoline supply, picks up a fine spray of gasoline. The mixture thus formed is in the proportion of about fifteen pounds of air to one pound of gasoline. After the mixture has been drawn into the cylinder, it is compressed and ignited as described in the preceding section.

The function of the carburetor, therefore, is to control the proportion of air and gasoline in the mixture that enters the cylinders. A mixture containing too much gasoline is said to be too "rich," whereas one containing too large a percentage of air is too "lean." A rich mixture provides a large amount of heat but burns too slowly for the engine to convert it into power. Therefore, the excess heat is lost in the exhaust gases and represents wasted fuel. A lean mixture, on the other hand, burns rapidly enough but, because the concentration of gasoline vapor is low, does not furnish as much heat as the engine could utilize efficiently. The result is a loss of power. For the sake of economy it is desirable to keep the mixture as lean as possible. However, this minimum will vary for different operating conditions. For example, when the engine is cold, it requires a somewhat richer mixture than when it has warmed up, since the cool cylinder walls absorb a greater percentage of the heat generated by combustion.

All carburetors are fitted with certain manual and automatic controls for adjusting the mixture to the optimum value. One such control is the "choke" valve which cuts off the supply of air and forces a heavier spray of gasoline into the manifold. A properly adjusted carburetor is an important factor in both the economy and the performance of an automobile. The throttle valve, which controls the speed of the engine, is mounted in the intake and manifold. The degree to which this valve is opened determines the amount of mixture drawn into the cylinder during the intake stroke. The speed of the engine increases as the amount of mixture increases; therefore, the throttle valve provides the means of controlling the speed of the engine and the speed of the vehicle that it is driving. As this valve is opened, the cylinders draw in a large charge of mixture, thus making the engine run faster.

The spark plug is the device that ignites the compressed mixture in the cylinder. It consists of two electrodes insulated

from each other by a porcelain sleeve and separated at the ends by a small gap. When the compression stroke has been completed, an electric spark is made to jump across this gap, thereby igniting the mixture of fuel and air in the cylinder and producing the explosion that transfers energy to the piston in the power stroke.

The electricity that supplies a spark to the spark plug comes originally from the automobile battery. When the ignition switch is turned on, current flows from the battery through the primary winding of an induction coil; and as the engine is set into motion by the starter, a mechanically operated switch called the "circuit breaker" opens the coil circuit each time that a spark is to occur. This induces a high voltage of about 20,000 volts in the secondary of the coil, and the distributor connects this high voltage with the cylinder that is ready to fire at that instant. The discharge then produces a spark across the proper plug, and the engine starts running. Once the cylinders start firing in their regular order, the operation of the ignition system is entirely automatic.

The combustion within the cylinders of a gasoline engine generates very high temperatures. Some of the heat is converted into mechanical work, as just explained. More than half of it, however, is absorbed in the iron walls of the cylinders and combustion chamber, and these parts get very hot. The combustion temperatures may be as high as 4500°F., almost twice the melting point of iron. It is essential, therefore, that a means be provided for getting rid of this waste heat. This is accomplished by the cooling and lubricating systems.

The cylinders and combustion chambers have hollow walls so that water can be circulated around them, this space being known as the water jacket. The water absorbs the excess heat from the cylinders, after which it rises through an opening in the top of the water jacket and flows downward through the radiator. The radiator is built out of a large number of very small tubes which are exposed to a stream of air. As the hot water flows through them, it is cooled by the surrounding air. The bottom of the radiator is connected to the lower part of the water jacket so that the cool water may return to absorb more heat. The water is made to circulate through this system

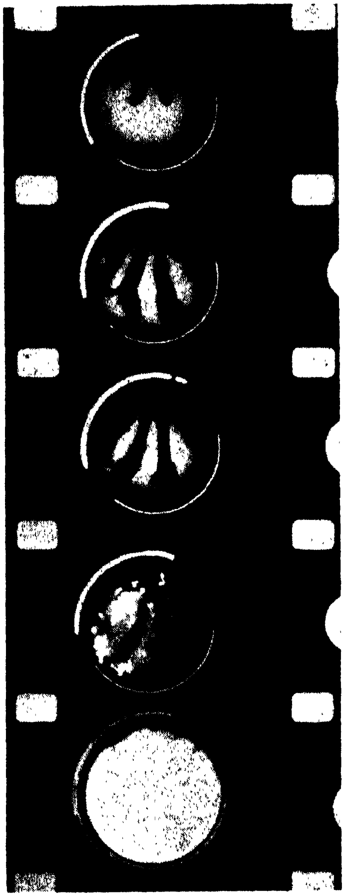
partly by convection currents set up by the unequal temperatures in the radiator and around the cylinder and partly by the action of a rotary pump which keeps the water in motion in one direction. Circulation of air through the radiator is maintained by a fan placed immediately behind it and driven by a belt connected to a pulley on the crankshaft.

Automotive engineers state that for every horsepower of useful work produced by a modern engine, one and a half horsepower must be carried away by the radiator. This is another way of telling us that the thermal efficiency of an automobile engine is low. Therefore, it must be cooled by circulating a heat-absorbing substance around the outside of the cylinder walls. Since water has such a large heat capacity, it is suitable for this purpose.

A cooling system does prevent melting of the metal of an internal-combustion engine, but a system for lubrication is necessary to prevent the moving parts from "burning out" or getting so hot from friction that they bind or stick. The cylinder walls and crankshaft bearings must be, therefore, constantly lubricated. For this purpose the housing around the crankshaft, called the crankcase, is partly filled with lubricating oil. Small oil trays are mounted under each connecting-rod bearing, and the connecting rods dip into these trays and splash oil on the cylinder walls and pistons as the crankshaft revolves. The trays are supplied with oil from the crankcase reservoir by means of an oil pump. The same pump also forces oil through pipes connecting all important bearings in the engine. The crankshaft is often made hollow, so that oil can be pumped through it to the connecting rods and main bearings. An oil-pressure gauge mounted on the instrument panel of the car serves as an indicator of oil-pump operation.

The oil used in automobile engines must have special qualifications. In the first place, it must retain its lubricating qualities at high temperatures. Although it must flow freely at all times, it cannot be so thin that it will not provide a film of lubrication between moving surfaces at high temperatures. The research laboratories of oil companies have made great progress during the last few years in the perfection of motor oil. This, in turn, has made it possible to build engines

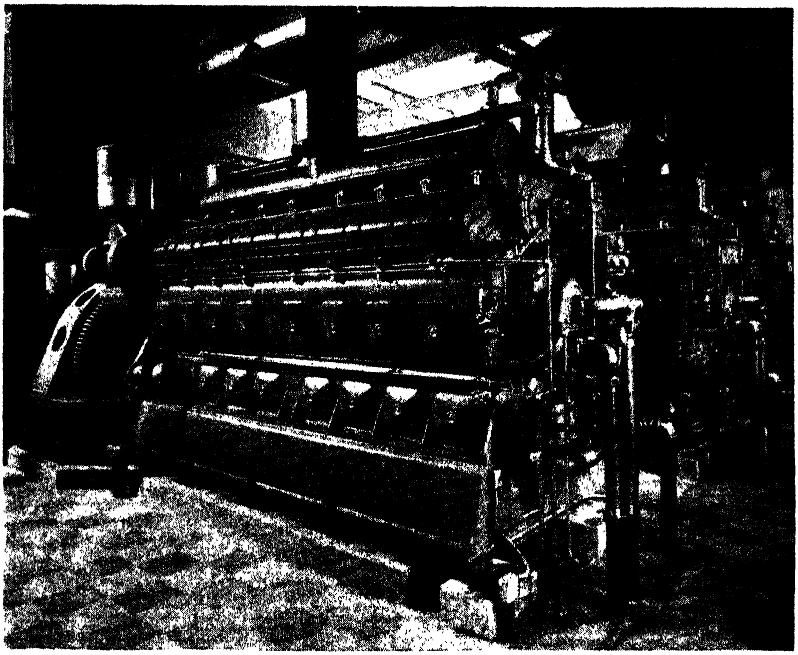
that operate at higher speeds and with less friction and wear than formerly.



These remarkable pictures of high-speed photography show the oil in the cylinder of a Diesel engine beginning to ignite at the top, increased burning in center pictures and extensive combustion in bottom picture. (Science Service photograph.)

An Engine without Spark Plugs

There is a type of internal-combustion engine that does not employ electric ignition and differs also from the automobile engine in other details of operation. A familiar example is the so-called "Diesel" engine, named after its inventor, Rudolph Diesel of Germany. It is designed to operate on cheap fuel, such



Diesel-engine installation for power and electricity in DuPont Building, Miami, Fla. (Wide World photograph.)

as crude oil or even fuel oil; in fact almost any hydrocarbon that can be rendered liquid will serve as a fuel.

Air is first drawn into the cylinder and rapidly compressed to about five hundred pounds per square inch. This rapid compression heats it. By means of an external source of pressure, fuel is then sprayed into the top of the cylinder, and the temperature of the compressed air is sufficient to ignite the fuel. Since the mixture of fuel and air takes place within the cylinder at a rather high temperature, no electrical system is necessary to ignite it. In this respect the Diesel is much simpler in design then the gasoline engine. It is not possible to vary either the amount of the mixture or the timing of the ignition over a very wide range, so the speed of the Diesel engine must be fairly constant for efficient operation.

Diesel engines are used in many kinds of service where large amounts of power are required. Some of the modern high-speed locomotives are equipped with Diesel engines that operate

electric generators, and the generators supply electric power to motors that drive the train. Also, some types of marine craft are Diesel powered. Since large Diesel engines are quite efficient, they are also coming into use for stationary power plants. The fact that they can use such cheap fuel makes them an economical source of power. They are quite heavy and do not as yet have the flexibility of control possessed by gasoline engines. They are constantly being improved, however; and since they can use such a wide variety of fuels, they will no doubt be an important source of power in the future.

How Much Work in One B.t.u. of Heat?

In the preceding chapter it was noted that heat energy is measured in terms of the calorie, or B.t.u. So far in this chapter we have seen how heat energy is converted into mechanical energy, but nothing has been said about how much mechanical energy can be produced by a given amount of heat. A knowledge of the quantitative relation between these two forms of energy is important because it enables us to measure how much work should be done by an engine when a given amount of fuel is consumed. The amount of work that the engine does compared with what it would do should there be no losses is a measure of its efficiency.

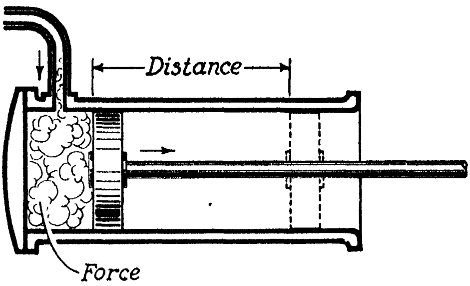
In order to make a quantitative comparison between heat and work, it is necessary to measure both of them in terms of the same unit. The easiest way to measure the work done by an engine is to take note of the force exerted by the piston and the distance through which it moves. The work done by each stroke of the piston is equal to the distance that the piston moves multiplied by the force that it exerts. If, for example, the piston moves two feet and exerts an average force of 150 pounds, the work done is found from the equation

$$\begin{aligned} \text{Work} &= \text{distance} \times \text{force} \\ &= 2 \text{ feet} \times 150 \text{ pounds} \\ &= 300 \text{ foot-pounds} \end{aligned}$$

It is obvious, therefore, that the logical unit for measuring mechanical work is the foot-pound, and we have seen that foot-pounds can be produced from heat energy. The question,

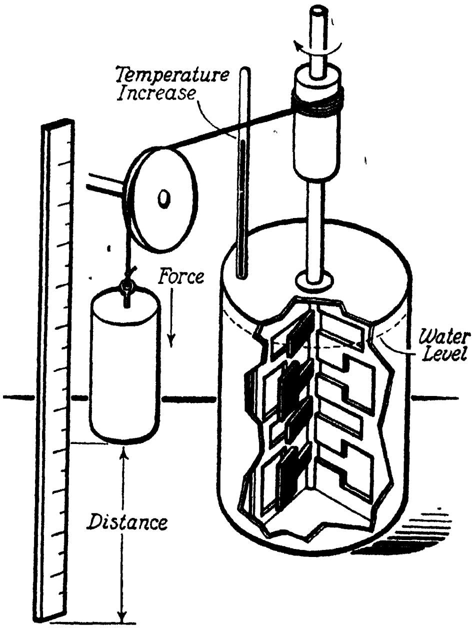
then, is how many foot-pounds can be produced from one B.t.u. It has been answered by accurate laboratory measurements.

The original experiments to determine the relationship between work and its equivalent in heat were made around 1850 by an Englishman named James Prescott Joule. A simplified form of his apparatus is shown in the upper drawing on the following page. The weight on the cord moves downward because of the action of gravity and rotates the stirring device in the vessel containing water. The water is warmed by the heat produced by this stirring motion. The vessel is well insulated so that none of the heat can escape, and an accurate measurement of its value may be made. The work done by the weight is equal to the product of the distance that it moves times the force that it exerts on the cord. The heat produced by the work is measured by the increase in temperature of the water. By carefully repeating the experiment a number of times, Joule found that 778 foot-pounds of work was required to produce one B.t.u. of heat in the water, a relationship that he called the "mechanical equivalent of heat." His measurements have been checked many times since by other experimenters, and the results have been repeatedly verified.

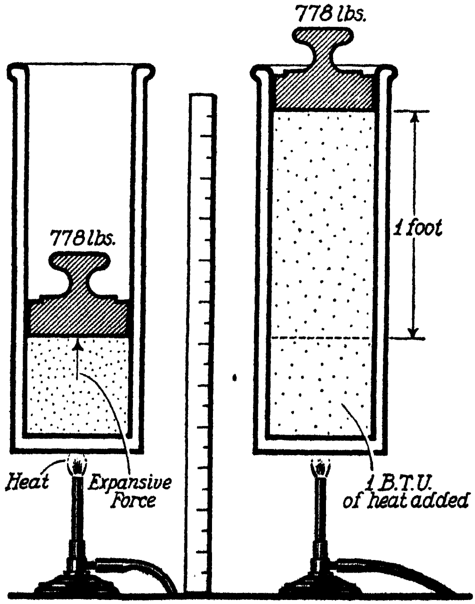


Work is measured by force × distance.

Thus we have an answer to the question of the relation between foot-pounds and B.t.u. Since the conversion of mechanical energy into heat is a reversible process, this relation also should apply to the production of work by a heat engine. The lower drawing illustrates the amount of work that would be produced by 1 B.t.u. in an ideal heat engine. If all the heat absorbed by the expanding gas in the cylinder were converted into work on the piston, each B.t.u. would raise the 778-pound weight one foot. This, of course, is a hypothetical device used merely for illustrating the mechanical equivalent of heat, as no such ideal engine exists in practice. The results of these experiments may be summed up in the statement that 1 B.t.u.



Work done by weight falling a definite distance generates a known amount of heat in the water.



One B.t.u. of heat can generate 778 foot-pounds of work.

of heat completely converted into mechanical energy will produce 778 foot-pounds of work, or that 778 foot-pounds of work when used exclusively to produce heat will give 1 B.t.u.

Getting the Most Work out of Heat

Most automobile drivers are interested in the mileage that can be obtained per gallon of gasoline, particularly if they have to pay for the gas, and economy of operation is a good talking point for any automobile salesman. Similarly, all power-plant superintendents try to obtain the maximum amount of power from each pound of fuel consumed. Economy of operation in this case is essential to profits or even, perhaps, to continuing in business. The extent to which the ambitions of either group will be realized depends upon the efficiency of the machine that they are operating.

During the history of heat engines a great deal of effort has been spent on developing methods for getting the most work out of a given amount of fuel. Laboratory tests show that the burning of a pound of coal produces about 13,000 B.t.u.; and a pound of gasoline about 20,000 B.t.u. How much of this potential heat energy can be converted into mechanical energy by a heat engine? The answer depends upon a number of factors, two of which are the size and type of engine. The efficiency for any particular case is concisely stated in the following relationship:

$$\text{Efficiency} = \frac{\text{mechanical-energy output}}{\text{heat-energy input}}$$

No engine ever performs according to the old proverb which states that "You get out of a thing what you put into it." This is a practical impossibility for mechanical devices, and it explains why no one has ever succeeded in building a perpetual-motion machine. Nevertheless, some types of engines are more efficient than others. From the standpoint of economy, it is desirable to make an engine operate as efficiently as possible, because even the best types are inherently wasteful.

To give an idea of how far we are from perfection in this field, some comparative figures are listed on the following page.

Type of Engine	Per Cent of Efficiency
Most efficient type of reciprocating steam engine	20
Most efficient steam turbine	40
Steam locomotive	10
Average automobile engine	25
Best Diesel engine	40

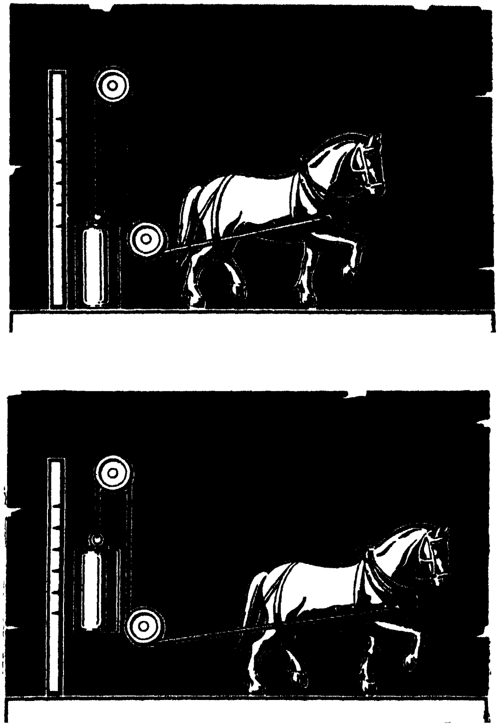
These figures show that even the best engines waste a large part of the heat supplied to them. This does not mean that designers of heat engines are outrageously incompetent. It does mean that there are certain practical limitations to the efficiency that can be obtained with a heat engine. One is set by the fact that a small interval of time is required for the expansion of the heated gas in the cylinder. During this interval some of the heat energy is conducted away by the cylinder walls and is lost. Another limitation is the fact that the expanding gas must be discharged through the exhaust at the end of the expansion stroke, although it still contains much heat energy. Loss of heat from either of these sources can be minimized to some extent by special design, but in neither case is it possible to eliminate all loss. For these reasons all heat engines have a relatively low efficiency.

Comparing Engines with Horses

When James Watt began to put his first steam engines on the market, he was confronted with the practical sales problem of showing that they could work more economically than horses. It was fairly easy to calculate how much the fuel for the engine cost, and every potential customer knew about how much it cost to feed a horse. The unknown quantity was how many horses could a given size of engine replace. Watt could measure how much mechanical energy his engines produced in a given length of time, but nobody at that time knew much about the number of foot-pounds of work that a horse could do during a day's work. In order to arrive at a basis of comparison, Watt made tests with the strongest dray horses available and concluded that a good horse could produce about 33,000 foot-pounds of work per minute. He named this quantity the "horsepower" and sold his engines with the specification that they would produce a certain number of horsepower when

running at full speed. This same unit is still in general use in Great Britain and the United States for specifying the power of engines. The power that an engine can develop depends upon its size and speed. Some of the largest steam turbines develop as high as 300,000 horsepower. For comparison, the average automobile engine can produce around 100 horsepower.

There is an important difference between the meaning of the words "power" and "work," although most laymen tend to use the two terms interchangeably. The amount of power required for a certain task depends entirely upon how fast the work must be done. A toy engine having one "mousepower" can do just as much work as the largest steam turbine ever built, but it would take it a much longer time.



In order to develop one horsepower, it is necessary for the horse to raise a 550-pound weight one foot each second.

The scientific meaning of power can be illustrated by the following example. Suppose that a ton of ashes is to be carried out of a basement to the sidewalk level and the depth of the basement is 12 feet. The work to be done is 12 feet $\times$ 2,000 pounds, or 24,000 foot-pounds. The power consumed in doing this work will depend on how fast it is done. Assuming that a man could carry 100 pounds of ashes per trip, he could do the job in twenty trips. If he possessed good endurance, he might be able to make the twenty trips in an hour. The power developed would be 24,000 foot-pounds divided by 60 minutes, or

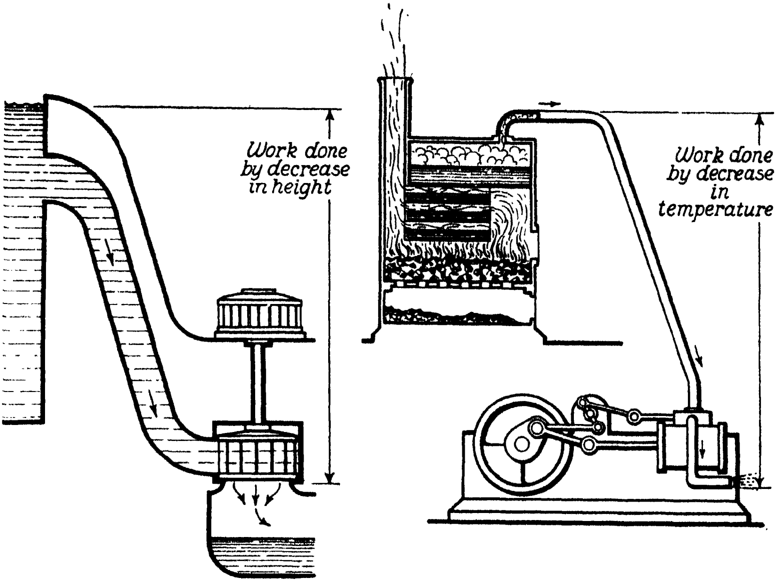
400 foot-pounds per minute. This is 400/33,000, or about 0.012 horsepower. If the entire ton of ashes is dumped on the freight elevator, it could be lifted to the street in about 10 seconds. The work done would still be 24,000 foot-pounds. The power developed, however, would be 24,000/10, or 2,400 foot-pounds per second. Since horsepower is equal to 550 foot-pounds per second, 2,400 foot-pounds per second would equal 2,400/550, or 4.36 horsepower. It is hoped that this example makes it clear that work is the total expenditure of energy, whereas power is a measurement of the rate at which this energy is expended.

Heat Cannot Flow Uphill

During the early history of the development of heat engines, little was known about the basic scientific principles involved in their operation. The progress that has been made in more recent times has come about largely as the result of the development of a branch of science known as thermodynamics. This is a long word and may seem to some readers a formidable one. Stated simply, it deals with the principles of converting heat into other forms of energy. There are two general laws of thermodynamics that apply to all heat engines, in fact to all conversions of heat energy. Their understanding provides an explanation for the behavior of any heat engine.

The first law states that when heat is converted into motion, the mechanical energy created is exactly equal to the heat energy that disappears. This is simply a restatement of the principle of the mechanical equivalent of heat discussed earlier in this chapter. Stated in popular language, it says that we cannot "get something for nothing," at least when working with heat and motion.

The second law tells us how much heat we can hope to get out of a heated object. The exact amount depends upon the temperature of the object and upon the temperature of its surroundings. Heat will flow out of a substance only if it can flow to a region of lower temperature. In other words, molecules will slow down if given a chance, but they will never speed up of their own accord. In this respect, the behavior of heat is analogous to the flow of a liquid. Water will not flow uphill



A heat engine is analogous to a water turbine in that in each case heat or water must flow downhill.

without the aid of outside energy; neither will heat flow "uphill" on the temperature scale.

This fact brings out the difference between the total heat of an object and its available heat. The total heat contained in any object is the amount of heat energy that could be extracted if it were cooled down to the point where its molecular motion ceased altogether. The available heat is the amount of heat energy that will flow out of the object as it cools down to the temperature of its surroundings. The total heat in a pound of iron, heated to 1000°F., is a fixed quantity but the available heat depends upon whether the iron is cooled in boiling water or in a refrigerator; *i.e.*, if the hot iron is cooled only to 212°F., less heat will be extracted then if it is cooled to 32°F.

Every pound of water in the ocean contains over 370 B.t.u. of heat. From a practical standpoint none of it is available because it will not flow out of the water. Lower temperatures could be created so that heat would flow out of sea water, but the energy necessary to maintain these temperatures would

exceed the heat energy obtained. Therefore the net gain in energy would be less than zero. This is the essence of the second law of thermodynamics. It says, in effect, that heat cannot be made to do useful work unless conditions are such that it can flow "downhill" in the temperature scale.

All the applications of heat both to scientific work and to engineering are based upon these two laws. They are as well established in physics as the law of gravity; nevertheless, uninformed inventors sometimes try to disregard them. Businessmen have occasionally invested large sums of money in projects that were impossible on the basis of thermodynamic principles.

REFERENCES FOR MORE EXTENDED READING

CLARK, W. H.: "Railroads and Rivers: The Story of Inland Transportation," L. C. Page & Co., Boston, 1939.

The development of transportation through these two mediums is well told and interestingly illustrated in this book.

BLACK, N. H., and H. N. DAVIS: "New Practical Physics," The Macmillan Company, New York, 1935, Chap. XV.

A well-illustrated discussion of the elementary principles of the steam and internal-combustion engine and their application.

BLACK, N. H.: "An Introductory Course in College Physics," The Macmillan Company, New York, 1935, Chap. XVII.

Those college students who wish a clear and concise treatment of steam boilers, steam engines, and gas engines will find this text a suitable reference.

SMITH, ALPHEUS W.: "The Elements of Physics," McGraw-Hill Book Company, Inc., New York, 1938, Chap. XXXI.

The above-noted chapter contains an exceptionally good explanation of the thermodynamic principles that apply to heat engines. The discussion includes isothermal and adiabatic changes and an explanation of the Carnot cycle. The treatment is accurate and not difficult to follow.

EWING, J. A.: "Thermodynamics for Engineers," Cambridge University Press, London, 1936, Chaps. III, V, VI.

Here is a very complete treatment of the thermodynamics of heat engines. It is recommended for those who wish to investigate the subject beyond the scope of an elementary course.

MARKS, LIONEL S., editor-in-chief: "Mechanical Engineer's Handbook," 3d ed., McGraw-Hill Book Company, Inc., New York, 1930, Sec. 9.

Although this handbook is intended for the engineer, a considerable amount of the discussion is understandable to anyone and provides a good explanation of how heat engines actually operate.

YOUNG, S. J., and R. W. J. PRYER: "The Testing of Internal Combustion Engines," English Universities Press, Ltd., London, 1936.

The practical problems of making tests for efficiency, horsepower, etc., are treated in this book. The subject matter would be of interest to anyone making a study of the operation of internal-combustion engines.

American Journal of Physics, published monthly by American Institute of Physics, New York.

The articles in this journal are related primarily to the instructional and cultural aspects of physics.

Power, published by McGraw-Hill Publishing Company, Inc., New York.

A monthly magazine devoted to practical engineering problems in the field of steam and Diesel engines and other prime movers. The subjects treated include power generation, transmission, application, and the attendant services in all industries.

Ewing Galloway.

10: NATURE'S INTANGIBLES

Being a Study of Energy in Wave Motion

IN FRANCONIA NOTCH of the White Mountains of New Hampshire a high, natural cliff forms an age-old profile of the human face. When a traveler through the narrow mountain pass stops at the right place and looks upward at the cliff, the face's remarkable contour is clearly visible against the sky. This huge rock of granite is known far and wide as "The Old Man of the Mountains." Old as the mountains themselves and born in temperatures so high that we have to tax our imagination

to consider them, this ancient rock might reveal much about the intangible nature of the universe.

It might say to us: "Yes, I am old and possessed with the knowledge of eons of time. I might tell you that the universe contains much more than you can actually see, hear, feel, and smell. That is because you have only human sight and other human senses; very limited perceptions, indeed. If you could have felt as I have, during the past millions of years, the wave-like pulses of power in sweeping surges coming out of the cosmos like fleeting spirits, you would be more conscious of what is going on around us.

"I have felt the daily surges of tremendous heat and light waves from the sun, since the ancient time when it was much hotter than it is now. I have felt these tremendous energies setting into terrific motion the very elements of which I am composed. These waves are much gentler today, producing only a pleasant tingle through my being. I can still feel on my surface the sun's ultraviolet waves. An entirely new wave motion I have experienced during the last human generation is of longer wave length. I understand that you call it radio and use it for communication.

"Another experience that I recall is the patter of the feet of many small animals and birds as they scrambled over my surface. These, too, produced waves that travel through me by causing mechanical vibrations within my substance. I realized that somehow they are different from the waves that come from outer space; however, they cause much the same effect as do the air waves that are produced when birds sing or wolves howl."

Thus might this huge formation of granite speak about various waves. Let us try to visualize at least some of the practical things that science has discovered about these unseen or often unnoticed energies.

Material Particles and Wave Energy

It is true that we have attempted to develop in preceding chapters the idea that all matter is built up of unit particles. We might recall in a brief review that the material substances of the universe are composed of molecules, which in turn are

made up of still smaller and more fundamental particles, the atoms. We have learned that chemical changes are reactions between these particles, and we have had an inkling that much of the energy employed by mankind is derived from chemical reactions. Furthermore, we have come to understand that the heat existing in matter is but a random movement of molecules; it is what might be called a "disorganized but continuing vibration of the particles constituting a substance." To our sense preceptions this seems, indeed, to be an atomistic world. However, a broader viewpoint must lead us to the study of wave motion to explain the bulk of the world's energy.

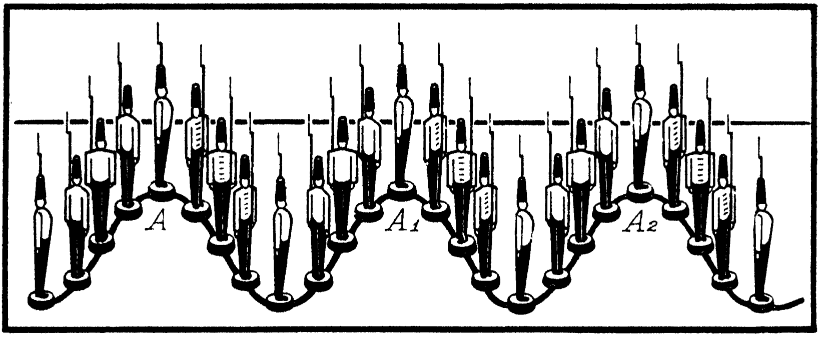
Some of the movements of the atoms or other particles of matter are of an orderly and rhythmic motion, producing within a substance what are called waves. Perhaps the most widely recognized types of waves of this character are water waves and sound waves and earthquake waves. Earthquake and water waves carry such enormous amounts of energy as to shake a skyscraper or demolish a flimsy building or lash to destruction houses along the coastline in the path of a hurricane. In addition to the types of waves that exist as rhythmic vibrations in material substances, another type of wave motion is exceedingly widespread in the universe. This consists of waves that travel through a vacuum and apparently do not involve the vibrations of particles at all. These waves make up what is called electromagnetic radiations, and they appear to be entirely divorced from matter. Such radiations are light waves, ultraviolet waves, radio waves, heat waves. The energy received from the sun consists primarily of heat and light waves, and it is the ultimate source of most of the energy stored and used on the earth. The energy available to man on the earth from sources entirely unrelated to radiant energy are dwarfed into insignificance when compared to the radiant energy received from the sun. Furthermore, evidence exists to indicate that even the smallest particles that constitute the ultimate structure of matter referred to above are but packets, or bundles, of wavelike energy. It might very well be argued, therefore, that the universe may consist entirely of basic pulsating phenomena rather than of particles. Although such an argument will not be presented in this discussion, it is a well-known fact that

Huge waves on the ocean as photographed from the deck of the S. S. Europa. (Photograph by Ewing Galloway.)

energy as wave motion constitutes a major proportion of the universe.

A Portrait of Wave Energy

Everyone has a mental picture of waves. One has only to stand by the seashore and observe the pounding of the water against the coastline to gain a visual impression of wave motion, as well as of the energy that it carries. Wave after wave lifts its crest, comes rolling in, dashes against the shore, and then recedes, to be followed by another. These waves come from far



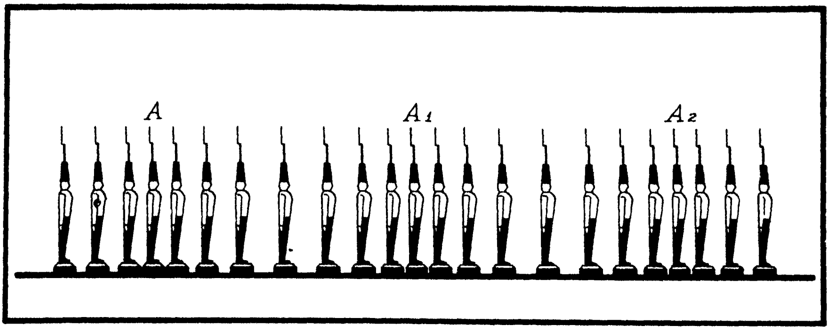
Representing a transverse wave.

out at sea, a foaming crest followed by a yawning trough. The water often rises and falls great distances, lifting with equal ease a floating log or an ocean liner. Water waves are waves in matter and consist of individual molecules of water moving as units in a rhythmic cadence.

Wave motion has certain attributes that are fundamental to its nature. To describe them, let us first mentally construct a working diagram. Suppose that placed on a perfectly smooth table top are a number of small tin soldiers to be used as units of movement and so constructed as to be able to move over the table in any direction, without friction. The tin soldiers are formed into one line facing front, one behind the other. They are connected with an elastic medium, such as a rubber band; but the medium possesses a compression strength as well as stretching strength. Now we are ready to begin.

Let us put the soldiers through a fundamental drill movements. The front one is moved a short distance to the right, then back to its original position, from there an equal distance to the left, and again returned to the starting point. The elastic band will cause the one immediately behind to execute this same movement at a slightly later instant. This, in turn, will produce a duplicate motion in the soldier behind it after the same interval of time. And so on down the line these movements will occur at equal intervals. Viewed from above, a sidewise displacement in the line will take place, moving along from right to left.

Now, if the right and left movements of the first soldier are repeated as soon as the first movement has been completed, the above-described motion of all will be repeated without any



Representing a longitudinal wave.

discontinuance, and a second displacement in the line will immediately follow the first. Should the movements of the first soldier be continued, other displacements will follow in continuous succession. The effect produced will be a wavy line, as shown in the left drawing, and the displacements constitute a series of waves traveling along the line. If the maximum displacement to the right is thought of as the crest, and the maximum displacement to the left as a trough, a wave motion is represented. Such a wave form in which a movement of the particles or medium producing the wave is perpendicular to the direction of motion of the wave is called a transverse wave. Light waves and other electromagnetic waves are transverse waves.

Another drill exercise can now be executed while we remember that the elastic connecting the soldiers has compression strength as well as stretching strength. The first soldier is moved a short distance to the front, then returned to the starting point, continued an equal distance to the rear, and back to the original position. The stretch of the elastic will draw the next soldier forward an instant later. Then the first soldier begins to move backward; a compression will be exerted on the second, and it, too, will begin to return an instant later than the first. Eventually the second soldier will execute the movements of the first. As the first soldier begins to move forward, the distance between it and the next increases slightly; and as it begins to move backward, the soldiers get closer together until their distance apart is somewhat less than it was before any movements took place. These same movements and

conditions will prevail between the second and third, the third and fourth, and so on down the line, in succession at short intervals later.

This motion of the soldiers, again viewed from above, will appear to be a separation of the soldiers proceeding from front to rear, followed by a jamming-together movement in the same direction. The separation may be called a "rarefaction"; and the jamming together, a "condensation." If these movements of the soldiers are repeated without interruption, the rarefactions and condensations will follow each other continuously. Again a wave motion will be seen traveling down the line from front to rear. In this case a movement of the particles or medium transmitting the wave is in the same line along which the wave is traveling, as represented in the drawing on the preceding page. Such a wave is called a longitudinal wave. Sound waves are of this character.

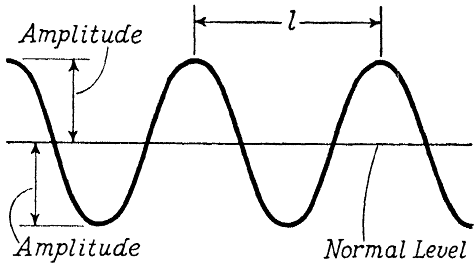
A third type of wave should be briefly considered. To illustrate, the soldiers are now made to execute their sidewise movements and at the same time to move slightly to the front and rear. The paths that they follow in these combined motions will not be straight lines but rather highly eccentric ellipses. The waves produced will be combinations of both transverse and longitudinal waves or somewhat rolling, like those on the surface of water. This roll of the water waves is well known by a swimmer who likes to swim in with the surf. He knows that the crest of the incoming wave will lift him and also carry him forward and that, as the wave passes, he will be lowered in the trough and retarded in his forward motion. The fine accomplishment of swimming with the surf, however, is to be able to remain on the crest of the wave and be continuously carried forward with it.

Measuring Waves

Now let us give attention to the relationship of wave length and frequency to the velocity of waves and, while doing so, point out some basic nomenclature.

The distance between any vibrating unit, or particle, in one complete wave and a similarly placed vibrating unit in the next complete wave is termed one wave length. To refer again to the

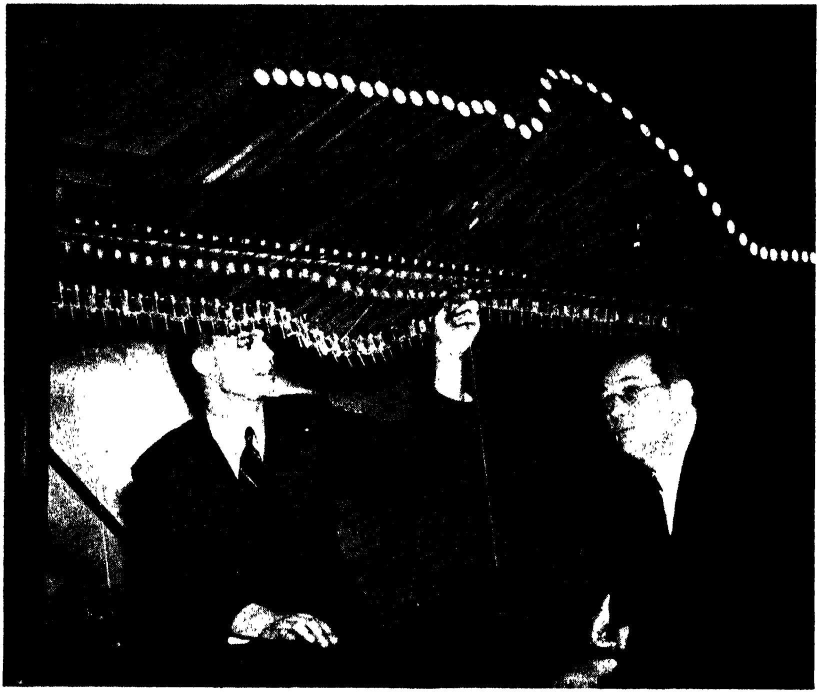
working diagram, as the displacements of the soldiers from side to side proceed with uniformity, the distance from one soldier (A) at the crest of one wave to another soldier (A_1) at the crest of the next wave represents a wave length. The same is true of the distance between any other two corresponding points on the wave form. This distance, or wave length, is usually designated by the letter l, and is so labeled in the accompanying drawing. It should be stated, also, that the distance from one condensation to the next in a longitudinal wave, or the distance from rarefaction to rarefaction in such a wave, is the wave length l.



The amplitude and wave length of a transverse wave.

The frequency of any wave is the number of complete waves, or cycles, passing a given point in one second of time, and it is always designated by the letter n. To refer once more to the soldier analogy, each time that the leading soldier executes a complete motion back and forth a new wave will be started. Each one of these complete back-and-forth movements is referred to as one vibration of the moving particles, and there will be the same number of waves per second as there are vibrations of the particles producing the waves. If five complete movements of the leading soldier are executed in one second, then the frequency of the wave motion is five, or $n = 5$. Waves are started, therefore, by vibrating particles and transmitted by vibrating particles or another vibrating medium.

Velocity of a wave motion is a term used to indicate the distance covered by one complete wave along its path in one second of time. Now suppose that the line of soldiers is ten feet long and the first wave travels from the front to the rear in one second of time. In this case the velocity of the wave motion is ten feet per second, the letter v being used to designate velocity. While the front wave of any particular wave is traveling along the line, five other waves will have been generated within the second's time, because the frequency of vibra-



A wave machine for simulating movements of electrical waves in motion invented by C. P. Wagner of the Westinghouse laboratories. (Courtesy of Westinghouse.)

tion is five per second. Accordingly, the entire line consists of five complete waves. Since they are of equal length, each wave will be two feet long. Here we begin to see the relationship between frequency, wave length, and velocity; *i.e.*, the velocity of a wave is equal to the wave length times the number of waves per second. This is expressed in a mathematical relationship as

$$v = nl \qquad \text{or} \qquad l = v/n$$

To illustrate the ease of determining one of these values when the other two are known, the figures given above may be substituted in the equation for determining, let us say, l.

$$l = 10/5 = 2 \text{ feet}$$

This relationship is fundamental to all wave motion and holds true for all forms of waves.

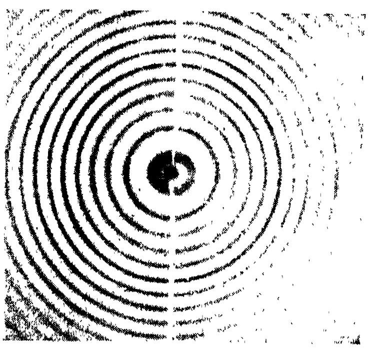
Another property of wave motion claims our attention at this point, namely, the amplitude of the wave. Amplitude is

the maximum displacement of the vibrating particles, or medium, from their position of equilibrium. In the case of a water wave, it is indicated by the height of the crest or the depth of the trough from the equilibrium of the water surface when there is no wave on it. The distance of the crest of the wave to the normal line is the same as the distance of the normal line to the bottom of the trough, as shown in the drawing on page 317, so that the amplitude may be measured from the mid-line to either the crest or the bottom of the trough. The amplitude is proportional to the intensity of the energy in the wave, a condition made obvious to us in water waves. In a calm sea the gentle waves may proceed unnoticed past a ship just as frequently as do the mighty mountains of water that toss it about on a stormy day. In the first case the gentle waves have only a small amplitude, indicating relatively little energy; in the other instance the waves from a storm are blown up to high crests and down to deep troughs, carrying with them such great energies as to rock and roll the largest ships. This fact, of the energy of a wave form being in proportion to the amplitude, is true of all other types of waves as well as of water waves.

Light waves radiate from a point source in expanding spheres.

Another fundamental attribute common to all wave motion is that waves radiate from a point source in expanding circles or spheres. Therefore, when a vibration, or oscillation, takes place at a point, the waves produced travel outward in all directions. A lighthouse or a single electric street lamp viewed at night from a distance shows the light spreading out in all directions. When an airplane soars overhead, the sounds of the motor are heard as it approaches, when it is directly overhead, and as it recedes into the distance. Sound waves are traveling outward from the plane in all directions. Also, if a stone is dropped into a body of water with a quiet surface,

water waves are seen to proceed from it in a series of concentric circles. Water waves are expanding circles, but light waves and sound waves are expanding spheres, if their source is very small, as it is in the case of the preceding examples.



A reproduction of actual waves radiating from a point source. The circles are wave fronts. (Photograph by Dr. C. A. Dyer.)

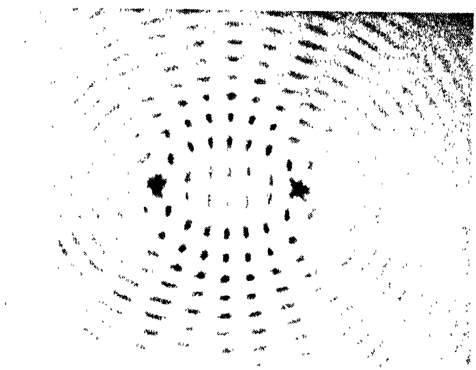
The direction of motion of a wave is from the source outward and is always at right angles to the wave front. Should an imaginary straight line be drawn from the source outward and perpendicular to the wave front, it would represent the direction of the wave; such a line is referred to as a "ray." Thus, the concentric circles or spheres, spoken of in the last paragraph, represent "wave fronts," and the radii of these circles represent the direction, or rays, of the waves.

Characteristics of Waves

To discuss and describe waves intelligently, it is necessary to know something about their characteristics, the most important of which are interference, reflection, and refraction. These characteristics account for many of the phenomena associated with energy as wave motion, such as beat notes in sound, amplitude modulation of radio waves by which the frequencies of sound are now transmitted by radio, the operation of mirrors and lenses in all optical instruments, and the production of the colors of the rainbow from white light.

The phenomenon of interference is the reinforcing and canceling action of two or more waves from different point sources passing and acting upon each other simultaneously. For example, when two sets of waves are carefully produced at two near-by points on a still surface of water in such manner that they are identical in all respects, interference is observed when they come together. There will be lines between the points where the waves have died out and other lines where the crests are higher and the troughs are deeper than those of either wave.

The two waves have had their energies canceled at certain places, and at other places their energies have been added to produce greater displacements. Furthermore, when two sounds of constant pitch are produced by two nearby sources of nearly the same frequency of vibration, a listener will hear louder and softer fluctuations of the sound, referred to as beat notes. The louder sounds are produced by the two waves adding their energies at certain places; the softer parts of the beat notes are produced by the waves canceling, or neutralizing, their energies at other places. Such is the phenomenon of interference.



Photograph of interference between similar waves from two point sources. (Photograph by Dr. C. A. Dyer.)

Perhaps an insight into why waves produce interference may be gained by noting the particular way in which the vibrating particles, or medium, transmitting a wave motion are related to each other. For this consideration let us refer again to the drawing on page 314 showing the displacement of a wave form, or the amplitude of a wave. In any wave motion all the particles at similar vibrating points, *e.g.*, those at successive crests, are at the same relative positions in the wave and are moving in the same direction at any instant. Such consecutive wave points are said to be "in phase" with each other. On the other hand, the particles at the crest and those at the bottom of the wave are at different relative positions and are moving in opposite directions at any given instant. These wave points are said to be "out of phase" with each other.

The two points, at the crest and at the bottom of the trough, are at opposite positions in the wave form, and particles at each point are moving in exactly the opposite direction. The opposite direction of motion will hold true at any instant of vibration, and these two points are said to be 180 degrees out

of phase. Likewise, particles at any other points in the wave form that are moving in different directions at any instant in their vibration are said to be out of phase, and the amount of out-of-phase relationship may vary from zero to 360 degrees. It should be kept in mind that any points at which the particles are moving in the same direction at the same instant are in phase, and any points at which they move in opposite directions are out of phase.

Now should two similar waves, such as the two water waves referred to above, meet each other so that they are 180 degrees out of phase, one of the waves will be attempting to move the vibrating particles of water upward toward the crest, and at the same instant the other wave will be attempting to move these same particles downward toward the trough. The result will be that the two energies will cancel each other at that place and the particles of water will not move at all; therefore, no wave will exist at that place. On the other hand, should the two waves meet so that they are in phase, the crest of one wave will be moving the water particles upward, and at the same instant the crest of the other wave also will be moving the same particles upward. The result will be that the two energies will move the water particles to higher crests and deeper troughs than would be true for either wave alone. At this place the waves reinforce each other.

The same conditions hold in the sound-waves illustration. When the two waves meet out of phase, the opposite direction of movement tends to cancel the vibration of the air particles transmitting the sound waves, and no sound (or only a low sound) is heard at that instant; but when the waves meet in phase, the movement of the particles is increased by the addition of the two energies in the same direction, so that the condensation is closer and the rarefaction is greater, and the sound louder. The increase and the decrease of sound constitute the beat notes heard.

What has been said of water waves and sound waves is also true of all other types of waves. In fact, interference as a result of the in-phase and out-of-phase relationship is true of all wave motion, but it is true of no other phenomena of nature. Interference in radio waves may be observed and measured

with the proper instruments, as likewise it may be observed in light waves.

When Thomas Young, the brilliant young English scientist, discovered in 1801 that light would produce wave interference, the wave theory of the nature of light was thought to be completely and infallibly established. Thus for a time was ended one of the most vigorous and extended controversies that has ever raged in the field of science. It centered around two hypotheses of the nature of light. One was that light consists of particles or corpuscles in motion; the other, that it is wavelike in character. The fact that this same uncertainty has bobbed up again in the twentieth century, with its exacting technique of scientific research, is indicative of the extreme complexity and elusiveness of the nature of light. But more about the nature of light in a later chapter.

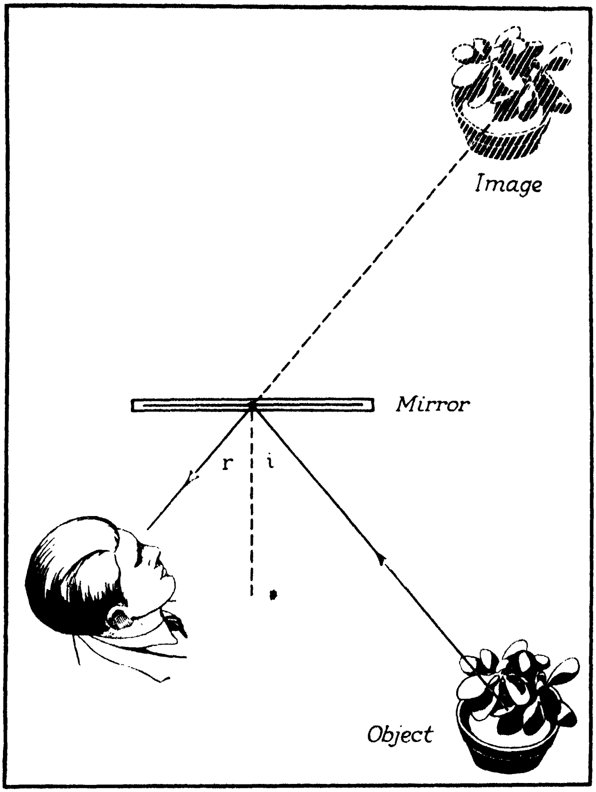
Another one of the important characteristics of wave motion is reflection; and most people probably have a general understanding of this phenomenon. Everyone has observed his image in a mirror; some have listened to echoes of their voices from a distant cliff; and a few may have noticed the water waves in a relatively quiet swimming pool strike the walls and turn back across the surface in the opposite direction. All these are examples of reflection of waves: the first a reflection of light waves; the second, of sound waves; and the third, of water waves.

Waves may be reflected from any surface smooth enough to turn them back without distorting their fronts too greatly. Just how smooth this surface needs to be will depend upon the size, or wave length, of the wave. Light waves have very short wave lengths, and a surface must be exceedingly smooth to reflect them without distortion. Sound waves, on the other hand, are much longer, and they may be reflected by a surface such as the face of a cliff or the wall of a building. However, the longer the wave lengths are the larger must be the surface in order to reflect the wave effectively.

If the reflecting surface is flat enough and large enough, the wave fronts will be reflected and travel as if they had originated at a point as far behind the reflecting surface as the source of the wave is in front of it. The place behind the reflecting surface from which the reflected waves appear to come is

This artistic photograph of an image in a mirror was possible because of the reflection of light waves by a smooth surface. (Photograph by Charles Heller, Philadelphia.)

considered the location of the image of the source producing the waves. Such an image is said to be "virtual," since no waves actually come from this location. With a plane, or flat, reflecting surface the virtual image will always be as far behind the surface as the source is in front of it. It is well known that one's image



The angle of incidence (i) made by a wave striking a reflecting surface is equal to the angle of reflection (r) of the reflected wave.

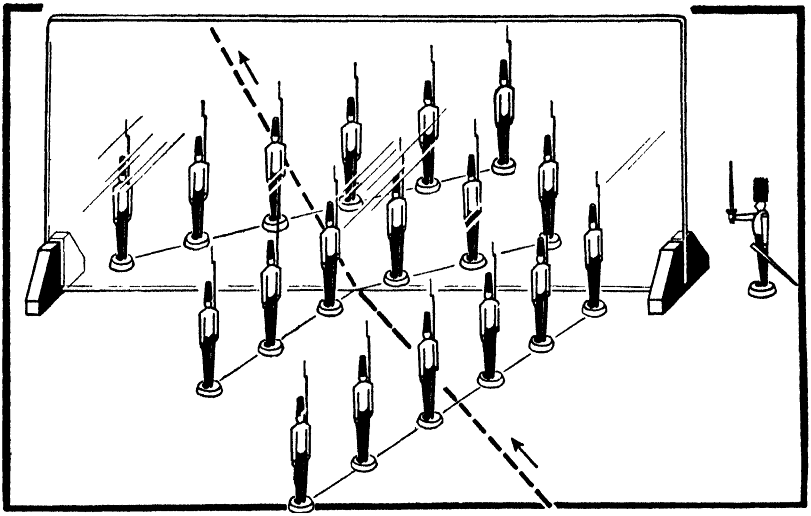
appears behind a plane mirror, and its distance appears to be as far behind the mirror as the person is in front of the mirror. That it is a virtual image may be surmised from the fact that one knows that no one of his exact likeness is standing behind the mirror, just as he also knows, when he listens to an echo of his voice, that no one is standing behind the distant cliff to mock him with his own words.

The particular angle with respect to the reflecting surface at which the image is formed will depend upon the angle at which the waves forming the image strike this surface. Suppose that a person is looking into a mirror and sees the image of an object at his right and to the rear, as shown in the accompanying drawing; the image will not only appear as far behind the mirror as the object is in front of it, but also it will be off to the side

by the same amount as is the object. The explanation of this is based upon a fundamental property of wave reflection, that the angle of incidence of the incoming ray is equal to the angle of reflection of the ray. These angles, respectively, are measured between a perpendicular line to the reflecting surface and a line representing the incoming ray and another line representing the reflected ray. They are usually designated *i* and *r*, as shown in the drawing. The reflected ray, extended (as shown by the dotted line) behind the plane mirror a distance equal to that of the incident ray, locates the image. The extension is a straight line and represents the direction from which the reflected ray appears to come, since the human eye has no power to distinguish any change of direction in a wave motion.

The other characteristic of wave motion is refraction, which means, in substance, the bending of the rays of the wave when it passes from a transmitting medium of one density into one of another density. Under ordinary circumstances the rays travel in straight lines, a condition particularly noticeable with light rays, since objects cast sharp shadows and since one cannot "see around a corner." However, under certain special circumstances light waves can be made to bend. This is the case when they pass into a medium of either greater or less density than that of the one in which they were traveling. To be more specific, the velocity of a wave motion changes when the wave enters a transmitting medium of a different density; and should the wave enter this different medium so that the rays strike the new medium obliquely, the rays would be bent.

This whole idea may be visualized by referring again to the diagram of the soldiers. This time they are arranged in a column, consisting, for example, of six abreast and one such line behind the other. They are set to moving forward, not, however, by an outside force exerted through a connecting elastic medium, as mentioned before, but each one independently and under his own power. In this instance the lines represent the wave motion itself rather than the medium transmitting the wave. So long as the soldiers all move with equal facility, their velocity will remain constant and their direction will be straight forward. However, assuming that they reach a locality or medium where there is added friction, the boundary of this medium being a



The lines of soldiers represent three wave fronts going from one medium into another. The medium beyond the screen is "harder going", the speed of travel is somewhat slower in this medium, and the direction is bent to the right.

straight line and parallel to the column front, the first six will enter this medium at the same instant, and their velocity will be equally reduced. Since they then begin to travel more slowly, the ones behind will close up somewhat until their velocity has been reduced accordingly by entering this medium. The direction of the column will remain the same but with a reduced velocity.

However, if the edge of this medium of added friction extends obliquely to the column front, the soldier, let us say, at the right end of the front line of six abreast will enter the denser medium first. His velocity is slowed down at once. The one next to him, traveling a little longer at the greater velocity before entering the new medium, will have moved a little farther in the same interval of time. The same is true of the third, the fourth, and along to the sixth. When all have entered the denser medium, the line has been bent around somewhat, the first soldier to enter acting as a moving pivot, or swing. The front line will then proceed at a reduced velocity but in a changed direction, as shown in the drawing. Somewhat the same condition is experienced in driving an automobile when one front wheel runs off the pavement. The velocity of the wheel tends

to slow down immediately so that the direction of the car is changed toward the wheel off the pavement. The other lines of the column of soldiers will follow the same procedure as the first line; and when the entire column has passed into the new medium, the column has slowed down and also its direction has changed at an angle to the original.

The lines of the soldiers abreast represent the fronts of the waves, and their line of march represents the ray. This bending of the ray by the passing from one density, or resistance to travel, into one of different density is refraction; and refraction is common to all types of wave motion. Refraction and reflection of waves also account for the focusing of rays under certain optical and other conditions, some of which will be considered in the following chapter.

Waves for Hearing

We have been taught from the time of our earliest nature-study classes that sounds are produced by waves in matter and that energy is transmitted along these waves. Most people have probably accepted this information on faith; and doubtless only a few have been interested and enterprising enough to verify the statements. A little thought given to well-known phenomena will, however, verify this teaching. Everyone has noticed that windows rattle when there is a sharp peal of thunder or a near-by explosion, and everyone has experienced a sensation of pain in the eardrums when a particularly loud sound was heard. The rattling of the windows is merely the result of their strong vibrations, caused by the intense condensations and rarefactions of the molecules of air beating against the panes. Such rattling is a response of the window to the sound waves, just as the sound that one hears is a response of the mechanism of his ears and his auditory nerves to sound waves.

The rattling of the windows is surely the result of continuing vibration. The condensations and rarefactions of the molecules of the air beating against the panes would set up such vibrations, but a sudden change in pressure against the window would produce a thud and not a continuing rattle. The rattle is always heard at the same time the distant sound is heard, indicating that the two are associated. The violence of the window rattle

as well as the loudness of the sound produced in one's ears are marks of the energy possessed by the sound waves that produced these results.

Such simple and easily observable conditions are hardly typical of the definiteness and accuracy with which sound-wave phenomena may be analyzed and measured. Sound phenomena constitute one of the best understood branches of science. A study of these principles and facts will easily enable everyone to have a better and more satisfying understanding of the waves constantly falling upon his ears. We may hear the rustle of the breeze through the trees, the song of a meadow lark, or the grinding of a Fifth Avenue bus and be able to recognize instantly the source of each without seeing it. This is because sound waves have a number of characteristics that enable them to carry the wealth of detail that makes it possible for us to distinguish one source from the other.

The first query that comes to the reader's mind, perhaps, is, What is sound, and how is it produced? From the remarks in the introductory paragraphs it might be surmised that sound is a series of longitudinal waves in matter of the proper frequencies; and from the standpoint of a purely physical definition, sound does consist of such waves. We do, however, associate with this physical phenomenon of wave motion the sensory stimuli and mental perceptions that result when the waves fall upon the ear. From the physical standpoint, sound is produced at any place and at any time where the proper longitudinal waves in matter are set in motion. This may be in a concert hall when an orchestra plays, or it may be in a forest when an acorn falls. In each case, longitudinal waves are produced in the surrounding air, and in a physical sense sound exists.

When the physiological reaction to the waves is included in the definition of sound, it means that the sound would be "heard" in the concert hall and probably not in the forest (at least by man). However, reading any such physiological response to sound waves into the definition of sound merely complicates the situation. A concert orchestra may be filling a hall with a variety of sound waves that will produce manifold sensory stimuli in the ears of one listener and at the same time move unheard past the ears of a deaf person. For our purposes here,

sound is to be considered as the physical phenomenon of longitudinal waves in matter.

Representing sound waves.

In a single statement the answer to how sound waves are produced is that they originate from any material or object that is vibrated in such manner as to bring about longitudinal waves of the proper frequencies. This is equally true whether the object be a vibrating tuning fork, a violin string, a bell, an organ pipe, the falling of water over Niagara, or the vibration of columns of air passing through the vocal cords. Vibrations may be produced in many ways such as by striking an object, by "bowing" it with a violin bow, by other mechanical means, and by electrical energy. To mention an interesting, but not common, example, a column of air in a properly chosen open pipe may be set into vibration by a small flame in such manner as to produce sound waves having a musical tone.

By the term "proper frequencies," referred to several times in the preceding paragraphs, is meant frequencies of vibrations that fall within the limits of frequencies to which the best human ear will respond. This range is from about thirty to approximately 18,000 vibrations per second. For most people any longitudinal waves of lower or greater frequency than these limits produce only a sensation of pain in the ears when the waves possess sufficient energy to affect the ears at all. Within recent years, however, microphones have been constructed that are sensitive to higher frequencies than the human ear can hear; and the upper limit of frequencies has been extended to about fifty thousand vibrations per second. These higher frequencies are usually referred to as supersonic sounds. Many studies of supersonic sounds indicate that certain living crea-

tures, such as varieties of insects as well as dogs, can hear frequencies far above the range of the human ear.

Another simple, and not uncommon, question regarding sound has to do with its natural transmission. Ordinarily the sounds that we hear are transmitted to our ears by the air. The air is a material medium, of course, consisting primarily of molecules of oxygen and nitrogen gases. These molecules are the particles that are set into vibration and thereby transmit the wave motion. It can be shown that other gases besides air transmit sound waves and that liquids and solids are even better carriers of sound than gases.

Most swimmers know that sound travels through water, apparently with greater ease and velocity than through air. To demonstrate this in the home or laboratory, a simple experiment may be carried out. Place a glass of water on a cigar box and fasten a cork to the stem of a tuning fork. When the tuning fork is set into vibration and the cork is placed on the surface of water, the sound will be heard as if it were coming from the box, showing that the vibrations of the fork have been transmitted through the water to the box; and the sounds from the box will be even louder than those from the fork reaching the ear direct through the air. Equally well known to apartment-house dwellers is the fact that sound travels through solids. This, too, may be easily demonstrated in the laboratory by taking a metal rod, placing one end of it against a door, and holding the stem of a vibrating fork against the other end. The sound of the fork seems to be coming from the door, and the metal rod serves as a sound carrier which transmits the vibrations of the fork to the door. The transmission of sound along the steel frames of modern buildings is now an acute problem of construction engineers.

It is to be concluded, therefore, that gases, liquids, and solids will transmit sounds. The velocity of sound in air is about eleven hundred feet per second; in water it is about four times as great; in steel it is approximately sixteen thousand feet per second. Sound waves are never transmitted through a vacuum, however. An explosion on the moon might be seen but never heard since the intervening space is a vacuum, even though

the explosion might be terrific enough to send sound waves that great distance through a material medium.

Some Characteristics of Sound

Sound has certain specific characteristics which it is in order to consider here, namely, pitch, interference to produce beats, quality, and resonance. All of them, with the possible exception of resonance, refer to the physiological reactions of the listener to the sound waves falling upon his ear; however, these subjective characteristics of sound depend in turn upon certain physical characteristics. By noting these relationships, it may be possible to gain a better understanding of how sound waves bring to the listener the wealth of information that they convey.

When we speak of a musical note as high or low, we generally mean its pitch. Stated somewhat differently, the pitch of a tone, as usually understood, is the place assigned to the tone in the musical scale. When we strike the keys of a piano in succession, we recognize the different tones produced as a difference in pitch. Likewise, when we strike two tuning forks that have different characteristics, we immediately say that the tones produced have a different pitch, and almost everyone, undoubtedly, is able to recognize a difference in pitch between his mother's and his father's voices.

The physical characteristic of sound waves that permits us to distinguish a difference in pitch of musical notes is the number of vibrations per second of the waves falling on the ear. The greater the frequency of the waves, or the frequency of the vibrating substance producing the waves, the higher will be the pitch of the musical note heard; the lower the frequency the lower will be the pitch; *i.e.*, the pitch of tone depends upon the number of vibrations per second of the sound waves.

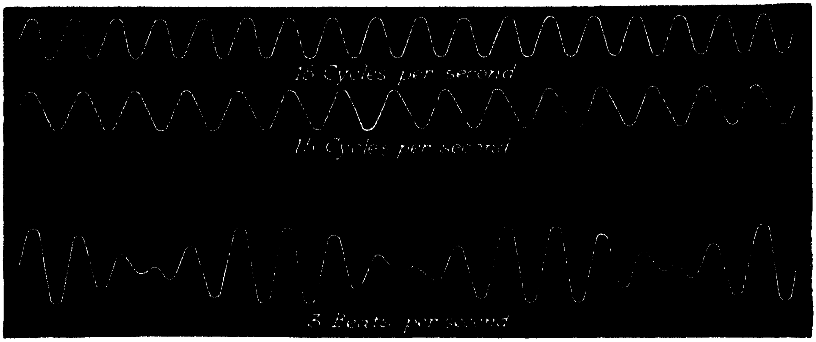
This relationship between pitch and frequency may be shown by a number of methods. One very interesting one is to produce a graphic representation of the sound waves from two tuning forks of different pitch by means of a cathode-ray oscillograph. This instrument produces a horizontally moving spot of light which vibrates up and down with the same frequency as the sound wave falling on a microphone properly connected to the

oscillograph. The moving spot of light "writes" a wave form which has all the features of the sound wave entering the microphone; and should the two waves produced by the two vibrating tuning forks be successively viewed in this manner, the wavy line resulting from the fork of higher pitch would be seen to have more waves but shorter wave length than the fork of lower pitch. A tuning fork of middle C will exhibit twice the number of waves as a tuning fork of the next lower C. To be more exact, the middle C fork (on the scientific scale) will have 256 waves per second, whereas the other fork will have 128 waves; consequently, the frequency of vibration of the first fork is 256, and that of the second is 128.

Many careful experiments with tuning forks and other vibrating sources of pure frequencies have led to the conclusion that if the number of vibrations per second is doubled, an increase of pitch of one octave is sensed; middle C of 256 vibrations is considered one octave above the next lower C of 128 vibrations. Likewise, another doubling of frequency, or 512, gives another octave increase, and still another doubling of frequency, or 1,024, gives an additional octave rise. The range of the standard piano is seven and one-third octaves, and the range of hearing of the human ear is about ten octaves. Within these limits the ear is sensitive to about 1,500 gradations of pitch. Some people with what is called "absolute pitch" are able to sense the change in pitch of a single vibration when the different notes are sounded at a fairly long interval of time.

For most practical purposes, then, pitch is a direct function of the frequency of the sound wave. However, when more detailed considerations of pitch are made, it is found that it depends slightly upon the loudness of the tone as well as the frequency. The relationship between pitch and loudness of a sound involves the response of the basilar membrane of the inner ear to the sound vibrations and will not be discussed here.

Another basic characteristic of sound is related to the principle of wave interference, mentioned earlier in this chapter. This is the production of beat notes. Should two sound sources, one of a frequency of 513 and the other of 510 vibrations per second, be set to vibrating in a large room, a listener standing



Two waves with a difference in frequency of three vibrations per second when combined will produce a beat note of three cycles per second as shown in the bottom drawing.

at a distance would hear a rise and fall in the loudness of the resulting sounds. This fluctuation in intensity would occur three times each second in the instance cited and would constitute a third frequency, which is the beat note. As a matter of fact, any two sounds with different values of pitch, if produced simultaneously, will beat together to produce a third sound. The frequency of the beat note is the difference in frequency between the two original sound waves.

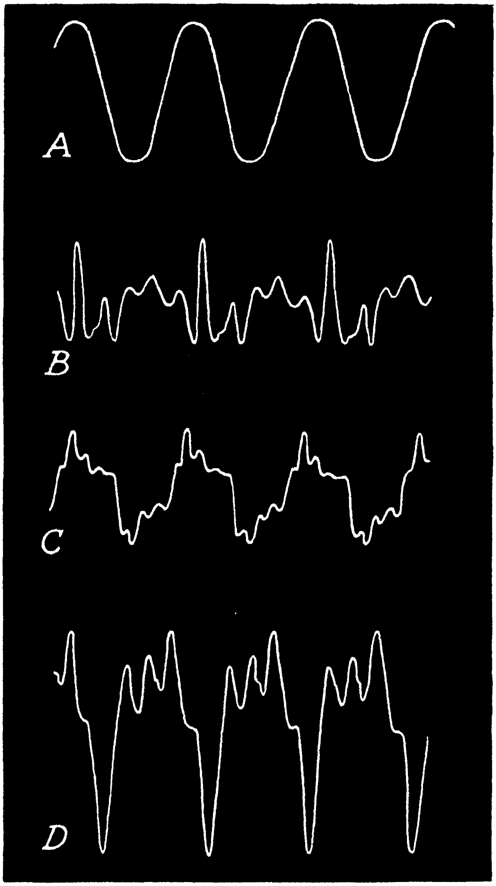
One can readily see that as the waves from the two sources travel to him, at one particular instant they will be in phase; *i.e.*, condensations from the two waves will reach his ear together, and so will the rarefactions. At that instant the two displacements will be added, and the result will be an increase in loudness. As the wave trains progress, the one of slightly shorter length will get out of step with the other. Presently, its condensations will arrive at the listener's ears simultaneously with the rarefactions of the longer wave. These two then will be out of phase, and the condensations of one will neutralize the rarefactions of the other. At this moment there will be no sound. Soon the shorter waves will have gained more on the others; again the condensations of the two will coincide, and the sound will have swelled again to its maximum strength. This will occur three times each second in the example given above, since the shorter waves are greater in number by three per second than the longer ones. The greater the difference in frequency of the two sounds the greater will be the frequency of the beat note produced in the new combined wave motion.

It is also possible to mix two sound waves whose frequencies are too high to be heard by the human ear and produce from them a note that is audible. Two sources producing sound waves with frequencies of 30,000 and 30,500 vibrations per second could not possibly be heard separately by the human ear, since these frequencies are beyond the range of audibility. Should they be set vibrating simultaneously, they would produce a beat note of a frequency of 500 vibrations per second. Thus, the beat note between the two very high, inaudible notes may be a definite pitch in the audible range. This process of producing beat notes between two different frequencies is known as heterodyning. By changing slightly the frequency of either one of the sources of high frequencies, it is possible to change the pitch of the beat note. The entire range of frequencies over the whole audible scale may be produced in this manner from two high-frequency inaudible sources. A musical instrument producing sounds in such manner was developed a few years ago. Known as the Theremin, it was for a time something of a musical sensation.

A third subjective characteristic of sound that has its explanation in the physical nature of sound waves is tone quality. Tone quality is what enables us to distinguish between notes of the same pitch and loudness when produced by different instruments or sung by different voices. We speak of the "velvet-noted" flute, the "humming" violin, the "voice with a smile" when we attempt to use words to express the tone quality of sounds. Let us see what physical characteristics of sound waves produce tone quality.

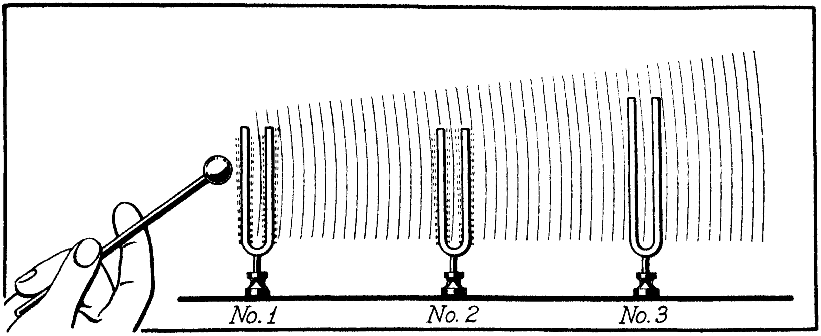
Quality is a function of the frequency of the sound wave, but it is a much more complex function than pitch. When the middle C on a piano (with the scientific scale) is struck, the string will vibrate with a frequency of 256 vibrations per second. This is its pitch, or fundamental tone. It may also vibrate simultaneously, but with lesser intensities, with frequencies of 512, 1,024, 2,048, and more vibrations per second. The sounds produced by these additional frequencies are called overtones. The sound wave emanating from the piano string, then, will not be a simple one but rather a complex one of the fundamental frequency with the frequencies of many overtones superimposed

on it. As a result, not only will the fundamental be heard but also many overtones of various frequencies and intensities. The exact relationship and loudness of the overtones with respect to the fundamental determine the tone quality of the sound heard.



Four waves of the same pitch but different tone quality: A, tuning fork; B, oboe; C, clarinet; D, saxophone. (Redrawn from Miller.)

The same condition is true of a vibrating bell or the human voice or most other musical instruments. All may have the same pitch, but the additional frequencies or overtones of each will be different. The bell will vibrate with a fundamental tone that may be matched by a violin string or a human voice, but its structure is such that its overtones will be different from those of the violin string, which in turn are entirely different from the overtones produced by the vibrating column of air from the vocal cords as it passes through the cavities of the mouth and nose. Likewise, the quality of different persons' voices is determined by the specific manner in which the overtones are affected by the particular configuration of mouth and nose cavities of the individuals. The complex wave forms of four different sounds of the same pitch but of different quality are shown in the accompanying diagram, wherein the fundamentals are represented by the large waves and the overtones are shown by the

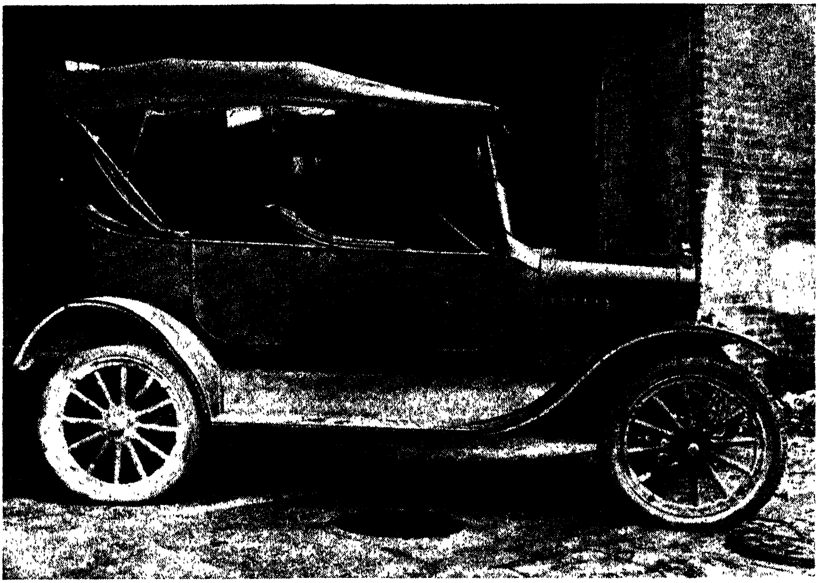


The phenomenon of resonance may be shown by the use of three tuning forks, two of which have the same frequency. Sound vibrations are produced by striking one of the forks with a hammer. The other fork of the same frequency will vibrate when the sound waves fall on it, but the third fork of a different frequency will not be set into vibration by these waves.

smaller and higher frequency waves superimposed upon the fundamentals.

A characteristic of the physical pulsations of sound waves is that such waves may set up "sympathetic" vibrations in objects other than the vibrating source of the waves. This phenomenon is resonance. Everyone knows from experience that it is easy to set a swing into long vibrations, making it go higher and higher by a succession of gentle pushes applied at just the right time and in the right direction, so as continually to increase its momentum. Mere random pushes, on the other hand, accomplish very little and may stop the swing altogether. In the same way sound waves may set up strong vibrations in an object if they are timed to correspond exactly to the natural frequency of vibration of the object, and thus resonance is produced.

A striking example of resonance is shown by holding down the pedal on a piano and singing a sustained loud note into the instrument. After the voice is silent, the strings may be heard reproducing the same tone with remarkable fidelity. Waves from the singing voice have set into vibration the strings having the same natural frequency as the singing voice. The vibrating strings, in turn, reproduce sounds of the same pitch. Two objects having the same natural period and connected by an elastic medium such as the air will both vibrate, therefore, when either one is set in motion. An open organ



In many old-fashioned cars, noises were produced at different speeds by headlights, hood, doors, or gears vibrating in resonance to the vibration of the car when their natural frequency was reached. (Science Service photograph.)

pipe of exactly the same natural frequency as a tuning fork may be made to produce its fundamental tone by feeble vibrations of the fork. The sound from the fork may be so low that it can scarcely be heard even a few feet away; but when the fork is held over the open pipe, the sound will be heard clearly throughout a large room. The air of the pipe has been set into resonant vibration by the waves from the tuning fork. The cumulative effects of the feeble impulses from the tuning fork repeated many times at regular intervals set up vibrations in the entire air column of the organ pipe whose natural period of vibration is the same as the impulses applied.

Resonance has its counterpart in mechanical vibrations other than sound. It has been said that a circus troupe of elephants marching in step across a bridge may shake it down. Regardless of whether such a remarkable effect has ever been produced, it is true that if the natural frequency of vibration of the bridge is the same as the ponderous thuds of the elephants' feet marching in unison, each successive impact will add to the intensity of vibration of the bridge. Common practice in march-

The remarkable acoustic effects on music were obtained somewhat accidentally in the architecture of Notre Dame Cathedral in Paris before the science of acoustics was fully understood. (Science Service photograph.)

ing a company of soldiers across a bridge is to have them march in "rout" order so that there will be no unison of footsteps which might possibly set it in vibration to the extent that crossing would be dangerous.

Making proper allowance for resonance is an important part of the design of various component parts of machines, where vibration may cause excessive wear and breakdown. If any one part of a machine has a natural frequency that corresponds to the frequency of a noise or other vibration produced by the machine, it will rattle excessively and wear itself as well as other parts with which it comes in contact. At the same time the machine will develop excessive noise. As an illustration of the latter point, it may be remembered by some that various parts on old-fashioned automobiles started to vibrate and to rattle as the speed of the car was increased or decreased to a given rate. This made the gears or other parts concerned wear out or shake apart faster, not to mention what it did to the occupants' nerves. Today, automobile manufacturers prevent these troubles by making sure that the natural frequency of the

various component parts will not lie within the range of frequencies of vibration that might be produced during normal operation of the car.

Resonance is an extremely important consideration in the architecture of auditoriums and in the design and manufacture of musical instruments. The shape of the air cavities of a violin or the sounding board of a piano has a marked effect upon the quality of the sound produced. Different parts of the violin cavities or piano sounding board are set into sympathetic vibration by certain resonant frequencies of the strings. These frequencies are thus amplified, while others are diminished or deadened, this being particularly true of the overtones of the vibrating strings. In such manner the quality of the sounds emanating from the instruments is noticeably affected. Likewise, in public auditoriums, cathedrals, or music rooms, if the space in any cavity of the room responds to one or more frequencies better than to others, a distortion of the sound waves will result. This may make it difficult to understand a speaker's words, or it may make music a jumble of sounds. The phenomenon of resonance has more influence upon the quality of one's voice than the vibrating vocal cords have. In fact the vocal cords serve primarily to set the column of air in the throat into vibration at given frequencies. The cavities of the mouth and nose act to strengthen or lessen the overtones by resonant vibration in such manner as to produce the specific quality associated with the individual's voice.

REFERENCES FOR MORE EXTENDED READING

JEANS, SIR JAMES: "Science and Music," The Macmillan Company, New York, 1937.

A versatile British scientist has turned his attention here to the principles of music. The book conveys precise information in a nontechnical way on the physical properties of sound, the physical mechanism of musical instruments, and the ways of making these instruments produce the most desirable musical sounds.

MILLS, JOHN: "A Fugue in Cycles and Bels," D. Van Nostrand Company, Inc., New York, 1935.

One of the foremost authorities in the sound engineering field has written an engaging book on how sound waves are handled.

SUTTON, R. M.: "Demonstration Experiment in Physics," McGraw-Hill Book Company, Inc., New York, 1938.

A comprehensive list of demonstration experiments to illustrate the principles and applications of physics. The demonstrations are so outlined and described as to enable the instructor or enterprising student to perform many experiments with relatively simple apparatus. Part II relates to wave motion and sound.

SMITH, A. W.: "The Elements of Physics," McGraw-Hill Book Company, Inc., New York, 1938.

Part II of this text presents a concise and well-written explanation of the phenomenon of wave motion and also of the production, characteristics, and transmission of sound.

PERKINS, H. A.: "College Physics," Prentice-Hall, Inc., New York, 1938, Part III.

The student wishing a thorough discussion of the theory of wave motion, fundamentals of sound, and operation of musical instruments will find no better reference than Professor Perkins' text.

MILLER, D. C.: "Sound Waves," The Macmillan Company, New York, 1937.

Without question the most analytical and authoritative study of the physical nature of musical sounds, the sound-producing characteristics of musical instruments, pressure effects in air of sound waves, and the propagation of sound waves from large guns. Written in a style that eliminates the necessity of technical knowledge for an understanding, and illustrated with a number of remarkable photographs.

Scientific American, published by Munn and Company, New York.

An excellent journal for the layman and scientist who desires to follow developments in science through well-written articles with a minimum of highly technical detail.

Journal of the Acoustical Society of America, published by the American Institute of Physics, New York.

The official news organ of the Acoustical Society contains professional and technical articles of interest to those working in the general field of acoustics as well as to the inquiring layman.

Bausch and Lomb Optical Co.

11: VISIBLE RADIATION

Light as One of the Six Divisions of the Electromagnetic Spectrum

ON MAR. 20, 1727, Sir Isaac Newton died at the age of eighty-five. Just before his death he is reported to have said, "I do not know what I may appear to the world, but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me."

This seems like a modest statement, indeed, to have been made by a man who had during his lifetime not only served as

Master of the Mint of the British government and been elected and reelected president of the Royal Society of London for twenty-five successive years but also had made some of the most important discoveries relating to the physical nature of the universe. Newton is probably best known for his formulation of the law of universal gravitation; however, his discoveries relating to the characteristics and properties of light were so extensive and fundamental as to give him a place among the greatest scientists of all time, even had these discoveries constituted his only contribution to increasing man's knowledge.

One of his first observations was that when a beam of sunlight is passed through a prism, it is separated into a spectrum of "very vivid and intense colors," and the illustration shown above is an artist's portrayal of Newton's discovery. He further discovered that when these colors are recombined they produce white light, like that received from the sun. This relationship between color and light is today considered an elementary concept, but in the seventeenth century the discovery was of primary significance. One of the troublesome questions in Newton's mind concerned the fundamental nature of light, and he was never able definitely to answer it. Perhaps this was one of the reasons that prompted him to make the statement quoted above.

Since Newton's time, it has been discovered that light is (in most of its characteristics) a form of wavelike radiant energy and that it belongs to a larger family of such energies, scientifically referred to as electromagnetic radiation. Further experimental investigation has revealed much information regarding the different forms of radiant energy. We know that the combined power of all the waves of all the oceans is completely dwarfed into insignificance in comparison with the tremendous and usually unseen energy of the electromagnetic spectrum.

Let us inquire into what has been discovered about some of the characteristics and properties of the different bands of the electromagnetic spectrum. In the present chapter we shall be concerned with the visible part of the spectrum, and in the next our attention will be directed to five other divisions of electromagnetic radiation.

The Nature of Radiant Energy

The fundamental nature of radiant energy is one of the most elusive problems of all nature. The idea of its electromagnetic character was first developed by James Clerk Maxwell, a British mathematician, in 1873, when he presented in a mathematical treatise his electromagnetic theory of light. Fifteen years later it was demonstrated experimentally by the German scientist Heinrich Hertz. Besides those waves which we know as light, Hertz had discovered another form of radiation belonging to the electromagnetic spectrum. These were the "electric," or short radio, waves produced by an electric current oscillating in an electric circuit. Since that time others have been added to the series. Now the electromagnetic spectrum is known to run the whole gamut of such radiations. Among the most commonly known forms are light, heat (infrared), and radio waves, and included also in this spectrum are the less familiar ultraviolet rays, X rays, and gamma rays. The only one of these six types of radiation that produces vision is light, as the five others are invisible, in that they do not stimulate sight in human eyes.

All these radiations, in addition to being electromagnetic in character, travel with the same velocity, are transmitted through empty space, and obey the same laws of refraction, reflection, and interference. They vary from each other fundamentally only in frequency or wave length and in their power of penetration through material substances. In wave length they range from several miles to billionths of an inch. The shorter ones have their wave length measured in a smaller unit than the inch, called an angstrom unit, which is one one-hundred millionth of a centimeter, or about one two-hundred millionth of an inch in length.

Before we begin a consideration of these types of radiation specifically, it is desirable to call attention to the fact that in view of recent scientific findings, which are not as yet complete or fully understood, the present theory of electromagnetic waves may have to be revised. Nineteenth century science had built up such a complete concept regarding electromagnetic radiation as to indicate that radiant energy was one of the best known

and best understood phenomena in all nature. Its speed had been accurately measured, and it was known to possess wave characteristics. The discovery by Young of the phenomenon of light interference, the brilliant mathematical treatise by Maxwell, and the demonstrations by Hertz seemed definitely to establish light and similar radiations as electromagnetic waves, transmitted through space in an intangible medium called the ether.

However, the beginning of the twentieth century ushered in new information which, with other discoveries since, indicates that light and other forms of radiant energy are something more than electromagnetic waves. They are now known to have two different aspects: one wavelike in character, the other corpuscular in nature, which means having the properties of moving particles. The exclusive wave nature of radiant energy began to break down when a most elaborate and accurate experiment performed to establish the existence of an ether failed to show any such medium. A medium, with the properties assigned to this hypothetical ether, is absolutely necessary to transmit light as waves through a vacuum. So far as we know at present, none exists.

Discovery of the photoelectric effect added other perplexities to the complete acceptance of the wave theory. The essence of this effect is that when light and some other forms of electromagnetic radiation fall on certain metals, the atoms of those metals have electrons ejected from them. Furthermore, when an electron is so ejected from an atom, it contains about seventeen thousand times as much energy as the atom would absorb in one second from a ray of light if the light were exclusively a continuous wave form. This means that it would require 17,000 seconds', or about five hours', exposure to light or other radiation before an electron could be ejected. However, an electron is ejected the instant that the light is turned on. To explain the photoelectric effect it is now suggested that light is not exclusively a continuous wave phenomenon but consists of unit drops of energy, called quanta, and that each separate quantum possesses sufficient energy to eject an electron the instant that it strikes the atom.

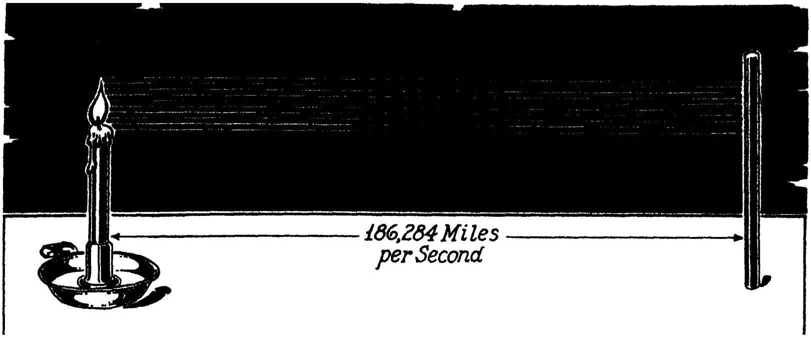
It would take us too far afield to recount the many other discoveries in the last twenty years that show that radiant

energy, particularly of the shorter wave lengths, is emitted and absorbed as quanta rather than as continuous wave motion. Suffice it to say that radiant energy manifests both wavelike and particle characteristics. As its wave length decreases, it tends to have more of the properties of drops of matter. In fact, radiation in certain special instances acts just like speeding particles. On the other hand, certain groups of particles, such as electrons, often have been observed to act like waves of energy. At present we are faced with this altogether disconcerting situation, one, quite frankly, that science to date has not been able to solve completely. More facts are needed. A further understanding of matter itself may disclose the solution of this apparent contradiction.

These puzzling new findings are characteristic of the remarkable and profound changes that are associated with the present era in the field of physics. The best that man can do at present is to measure the effects of radiant energy when it is absorbed and to study the nature of its sources. What radiant energy really is in its flight through space, what its exact nature is, is still a mystery to be solved. However, most of the characteristics of electromagnetic radiation in motion can be explained on the basis of the fact that it consists of high-speed waves. Regardless of the unknown nature of these radiations, it should be clearly understood by the reader that many of their wave properties are well known and that the laws that they obey are as firmly proved as is the law of gravity. Such is particularly true of their wave characteristics, and these are the characteristics that will be considered in this text.

Visible-light Radiation

The most common form of electromagnetic radiation is visible light. There is no mystery about its sure existence, what it does, and how it acts. Even though it is one of the most intangible things in the universe, it has many of the properties of the well-known physical waves discussed in the preceding chapter. It is produced by the vibration of electrons in heated atoms, and its silent passage through space has brought us all the knowledge that we have of objects outside the earth.



The speed of light has been accurately measured as 186,284 miles per second.

Light travels at a constant and terrific speed in a vacuum and only a little less rapidly through transparent matter. The velocity of light in a vacuum is 186,284 miles per second. It seems, from casual examination and observation, to travel instantaneously. It does, however, under very careful measurement, exhibit a finite velocity, which is the same as the velocity of all other kinds of electromagnetic radiation. In the metric system this speed is about 300 million meters per second. Great as this is, it has been measured with an accuracy of 99.99995 per cent.

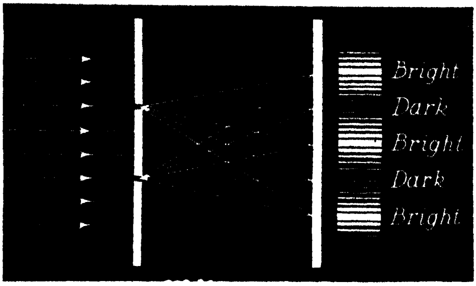
The high degree of precision is the result of nearly fifty years of work in this field by one of America's greatest scientists, Albert A. Michelson. Using an instrument of his own development, he continually repeated his measurements in order to get them each time a little more accurate. At the time of his death he was engaged in the most elaborate and extensive attempt to measure this last mile of accuracy, and the first figure given above is the result of his data. One may question the value and necessity of a lifetime's devotion to securing such accuracy. Its significance lies in the fact that the speed of light is the maximum speed attainable in the universe, and it represents one of the very few absolute constants of nature. Upon the accurate measurement of the velocity of light depends our ability to measure time accurately as well as accurately to measure distance in its most fundamental aspects.

Light waves are extremely short, with a range in wave length between 7,800 and 4,000 angstroms according to the

Albert A. Michelson (1852–1931), professor of physics at the University of Chicago and Nobel prize winner in 1907, was one of America's greatest scientists. (From Black and Davis, "Elementary Practical Physics," The Macmillan Company.)

color of the light. For instance, green-yellow light, which the human eye sees best, consists of waves of about 5,500 angstroms in wave length. This color lies near the center of the visible spectrum.

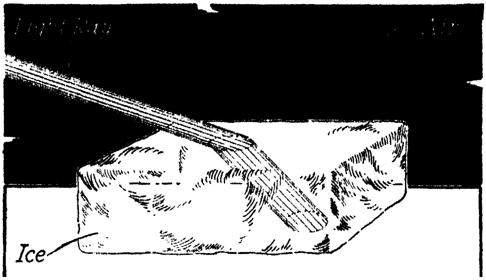
That light does act like waves and has a measurable wave length is proved by the phenomenon of interference. Light waves from a common source when sent through narrow slits may be made to "beat" with one another and produce alternate areas of light and dark on a screen, as represented in the



A demonstration in wave interference where light from a common source passes through two tiny slits onto a screen. In some places it cancels out, leaving a dark area, and in other places it produces a brighter area, as represented at the right in the drawing.

accompanying drawing. The idea of placing two slits in front of one light source is in reality to provide two sources of light, one at each slit, that have exactly the same wave and energy characteristics. When such apparatus is properly set up and operated, the waves that come through the two slits are found to add their energies in some areas and to cancel them in others to produce light and dark areas, respectively, as shown in the illustration. The areas in which the crests of the waves from one slit meet the troughs of the waves from the other slit will be dark; those in which the crests of waves from both sources meet will be light. It is possible to determine the wave length of light by measuring the distance from one dark area to the next and considering the relationship of this distance to the wave length of the waves producing the interference. When light of a single wave length, *i.e.*, light of a pure color, is used, that particular wave length may be determined. By changing the color of the light, it is possible to measure the wave lengths of all the different colors. Red light has wave lengths of approximately 7,500 angstroms; orange, 6,300; yellow, 5,800; green, 5,400; blue, 4,800; indigo, 4,400; and violet, 4,000 angstroms.

Refraction is another phenomenon characteristic of light waves. Most people know that light rays travel in a straight line when passing through a medium of uniform density, such as the air. When the ray, which is simply the direction of travel of a small section of the wave, passes from a transparent medium of one optical density into a medium of a different optical density at an angle other than ninety degrees, the ray is bent.



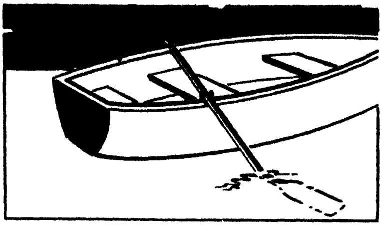
A beam of light is bent upon passing from air into ice.

By optical density is meant the characteristics of a transparent substance that determine the velocity of light in passing through that substance. Probably everyone has noticed some example of the refraction of light. A common illustration is the apparent bending of a spoon placed in a cup of transparent liquid; also, the apparent bending of an oar when dipped at an angle into water. In each of the illustrations cited, the apparent bending is in such direction as to make the part of the object beneath the liquid seem nearer the top of the liquid than it would if no bending were obvious.

The explanation of such bending is based upon the slowing down of the velocity of light waves as they pass into a medium of increased optical density, as explained in connection with the marching soldiers in the preceding chapter. If the light ray strikes the new medium at ninety degrees to all parts of the wave front, the entire wave is slowed down at the same instant, and it continues to travel in its original direction but at a reduced velocity. However, should the ray strike the new medium at an angle so that one part of the wave front enters before the other parts, the first part would be slowed down first, and this would cause the direction of the wave to bend, as illustrated in the drawing in which a ray of light is represented passing from the air into a block of ice. The ray is bent downward, as is to be expected, since the underneath sides of the wave fronts strike the denser medium first.

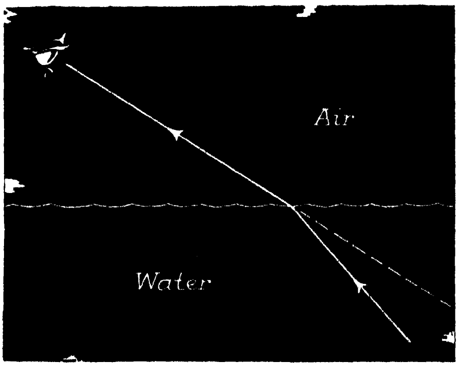
The reason why the oar seems to be bent upward when a part of it is dipped into water is that the reflected light rays from the oar upon emerging from the water into the air are bent in the same direction as might be represented in the

ice drawing, should the light ray be considered as coming from the ice into air. A person would see the part of the oar beneath the water by the light rays entering the eye after they have emerged from the water and passed through the air. Since the eye is not able to perceive any bend in the light rays entering it, the rays would appear to be coming in a straight line from points above where the oar actually exists, as shown in the simplified drawing of one light ray; therefore, the oar appears to be bent upward so as to coincide with these points.



The apparent bending of an oar in water is produced by refraction.

It probably is easy to understand that the amount of bending produced in a light ray upon entering a medium of different density at any given angle is determined by the amount by which the velocity of light is changed. If the velocity is reduced considerably, the ray will be bent more than when the reduction is less. It is possible, therefore, by measuring the velocity at which light travels in that medium and comparing it with the velocity of light in air, to determine exactly how much bending will be experienced by a ray of light upon entering any transparent medium. (Actually it should be compared with the velocity of light in a vacuum. However, for practical purposes, the velocity of light in air is used, since most of our experience with light involves its velocity in air, and since its velocity in air is only slightly less than its velocity in a vacuum.) The ratio of the velocity of light in air to its velocity in any other transparent medium is a measure of what is called the refractive index of that medium, which expresses the degree of



A person would see the oar at the point from which the light appears to come, as represented by the dotted line.

bending experienced by the light rays in going from air into the new medium or in coming out of it. The refractive index may be concisely expressed as follows:

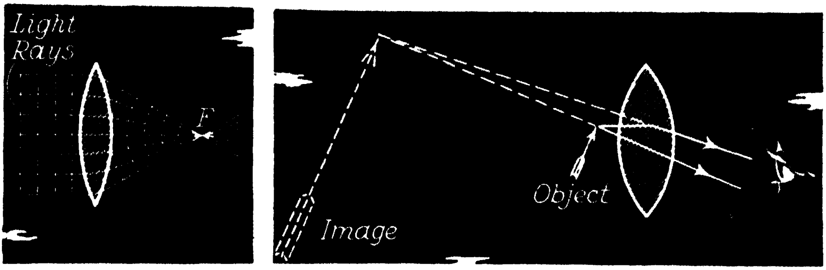
$$\text{Index of refraction} = \frac{\text{velocity in air}}{\text{velocity in other medium}}$$

The figure thus obtained for any transparent substance is a convenient method of expressing how light will behave in passing through that substance. For example, the velocity of light in water has been measured and found to be about 140,000 miles per second. The index of refraction of water, therefore, is 186,284 divided by 140,000, or 1.33. The index of refraction of all other transparent substances has been measured. That for ordinary crown glass is 1.43; that for special flint glass is 1.68. The diamond has a refractive index of 2.47, and this helps to account for the diamond's unique property of refracting light so as to give the beautiful play of colors for which it is famous.

It is possible to take any transparent substance and, by grinding its surfaces into various shapes, make light rays bend at any desired angle upon entering it and again passing into the air. If the surface is ground into a shape other than a plane surface, it is impossible for an entire wave front of any appreciable dimension to enter at the same instant, and accordingly some bending of the ray must occur. Transparent substances with their surfaces ground to a smooth curve constitute lenses, and the variety of lenses that can be made is limited only by the different kinds of transparent substances available and man's desire and ability to shape their surfaces into various curves. Cameras, telescopes, microscopes, searchlights, eyeglasses, opera glasses, even our eyes themselves depend for operation upon the refraction of light by lenses.

Lenses Produce Optical Illusions

Since the eye is unable to perceive the bends in the rays produced by refraction as the rays pass through lenses, various illusions of enlargement, reduction, or distortion of the object viewed through a lens may be created. Just what illusion is produced will depend upon the shape, and to a certain extent the substance, of the lens. There are two fundamental types of

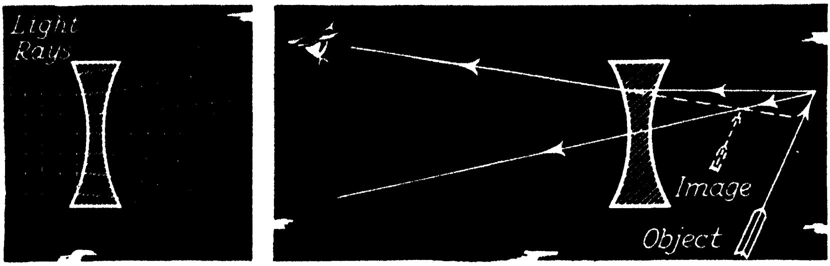


Bending of light rays by a convex lens (left), and a representation of why objects viewed through such a lens appear larger (right).

lens as determined by their shape, known as convex lenses and concave lenses.

Convex lenses have both surfaces curved outward so as to be thickest at the center. A beam of light passing through this type of lens is bent, or "condensed," to a focal point at the opposite side of the lens from the source, as illustrated in the drawing. If an examination is made of what happens to the wave fronts as represented by the broken lines in the drawing, it will be obvious why the direction of the rays is bent so as to condense, or focus, the rays at *F*. Such a lens is used for magnification. Objects viewed through it appear to be enlarged and brought nearer the observer, and this is so because we see them at the place from which the rays appear to come without being bent, as illustrated by the dotted lines in the adjacent drawing. The greater the curvature of such a lens the shorter will be the focal length and the greater the magnification. A modification of the convex is the planoconvex lens, a lens that is flat on one side and curved outward on the other. Its light-bending properties and magnification are the same as those of the convex lens, only less pronounced.

Concave lenses have both surfaces curves inward and are thickest at the edge. A beam of light passing through a concave lens is bent outward, or "scattered," on the opposite side from the source, as illustrated in the next drawing. The wave fronts are again represented by broken lines, and their shape after passing through the lens shows why the rays are scattered. Objects viewed through a concave lens appear to be reduced in size, as may be apparent by examining the drawing to represent this condition. Such a lens is called a reducing glass. The greater



Bending of light rays by a concave lens (left), and a representation of why objects viewed through such a lens appear smaller (right).

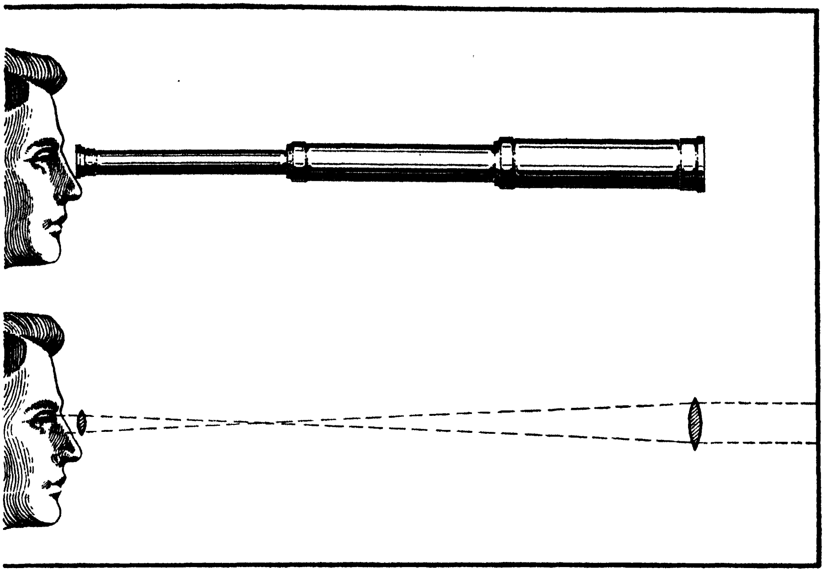
the curvature of the lens the greater will be the scattering of the rays, and the greater will appear the reduction in size of the object viewed. The planoconcave lens is flat on one side and curved inward on the other. It produces about the same light-scattering effects as the concave lens, except to a lesser degree.

Curved mirrors also may be made to act like lenses in focusing light rays. However, mirrors of different curvature focus or scatter the light in a manner just the opposite from that of lenses. A convex mirror spreads the rays of light as they are reflected in the direction from which they came; a concave mirror condenses the light rays so that they come to a focus in front of the mirror. The disproportionate images that one sometimes sees of himself in certain mirrors at amusement parks are produced by odd combinations of these two types of curvature which give unusual magnifications and reductions, and thus bring about the strange effects.

Some Basic Optical Instruments

The refraction of light waves through lenses or their reflection from mirrors permits the design and manufacture of many optical instruments. Such instruments have provided us with much of our information regarding the nature of the universe and afforded us most of our pictorial records, and they enable us in many ways to see objects more clearly than with the unaided eye. A brief study of a few of these instruments is worthy of our time at this point.

Refracting telescopes consist of an adjustable lightproof tube with a set of lenses mounted at each end. A large image-forming



A refracting telescope contains a large objective at the front end for collecting light and a smaller eyepiece for magnifying the image.

convex lens, known as the objective, is placed at the front end of the tube, and the other and much smaller lens, called the eyepiece, is mounted behind the focus of the objective at the other end. The large objective gathers as much light as it will intercept from a distant and usually dimly lighted object and focuses it to a small image of the object. The purpose in having a large objective is to gather sufficient light to make the image as bright as possible. The eyepiece needs to be only a small convex lens so as to magnify the image properly and focus it on the eye or a photographic plate. The distance between the two lenses is made variable by the adjustable tube for focusing purposes when viewing objects of greater or less distance. The tube also excludes all light except that entering through the objective lens and thus prevents unwanted light from blurring the image.

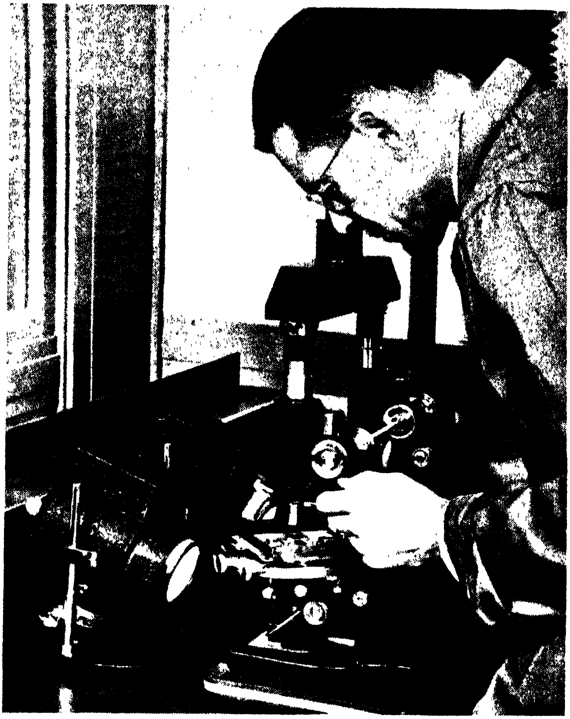
Telescopes of the reflecting type use curved mirrors rather than lenses for objectives. The action of the mirror on the light from the object is the same as that of the lens, which is to focus it so as to form an image. It is necessary to remember,

however, that the mirror must be concave rather than convex in order to produce such an image. The eyepiece in a reflecting telescope serves the same purpose as it does in a refracting instrument, to magnify and view the image formed by the objective. Most of the very large telescopes have mirrors rather than lenses for objectives. The main reason is that it is much more practical to make the finest type objectives in large sizes as curved mirrors rather than as lenses; and, of course, such large objectives must be as nearly perfect as possible in order to justify the expense of making and using them at all.

Field glasses and opera glasses are essentially two small refracting telescopes placed side by side so as to provide one for each eye. Viewing the object by each of the two eyes separately permits a person to perceive the depth in the object and gives us what is called binocular vision. No other way is known at present to achieve binocular vision. Field glasses and opera glasses, by their very construction, produce in each eye a separate image of the object viewed and thereby afford enlarged binocular vision.

The instruments just described—telescopes, field glasses, and opera glasses—are used to examine distant objects by forming as bright an image as possible and then magnifying it as much as is necessary. They create the illusion of bringing the objects nearer.

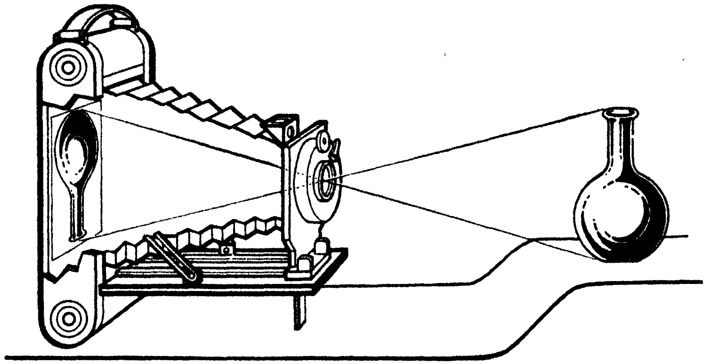
The microscope has certain features in common with the refracting telescope, in that it consists of an objective lens and an eyepiece, and these are mounted at the opposite ends of a lightproof tube the length of which is adjustable. Furthermore, the objective lens forms an image of the object, and the eyepiece is used to view this image. As everyone knows, the microscope is used to view very small objects that the unaided eye cannot see. The design of the objective must therefore be entirely different from that of the telescope. It must be a very small convex lens of great curvature, rather than a large one of little curvature, as is true of the telescope. The greater the curvature of the lens the more the rays of light will be bent in passing through the lens, as was noted earlier in this chapter. The image produced by the microscope objective will be, therefore,



A new and powerful comparison microscope, that is really two instruments in one, being used by a Federal Bureau of Investigation scientist. The objective lenses are focused just above the object on the glass slide which is itself strongly illuminated by the lamps seen at the left. (Science Service photograph.)

much larger than the object, and the best microscope objectives will produce clear magnifications of over 100 diameters. This great magnification necessitates that the object be brilliantly illuminated so that the enlarged image will still be bright enough for further magnifying and viewing by the eyepiece. A strong light must be concentrated on any object to be viewed with a high-power microscope. The eyepiece of the microscope includes lenses that are larger than those of the objective; and it is used, as in the case of the telescope, to magnify the image formed by the objective. In fact, it is not impossible to use the same eyepiece interchangeably on the two instruments.

Binocular microscopes are now made to give perspective and depth to the microscopic image, very much as in the case of



The camera is an optical instrument in which a lens bends the light rays from an object to form an image on a sensitive plate.

field glasses. Binocular microscopes, even though they may be of lower magnifying power than ordinary microscopes, are more useful in the study of crystal formations and in any instance where the shape of the microscopic object or the depth in the object is the important thing to be observed.

It is interesting to note that these same general instruments, the telescope and the microscope, are used for examination and study of the largest things in the universe, the stars and nebulae, and also the smallest visible objects in the universe, metal structure or structure of living organisms. Each of these instruments uses condensing lenses or mirrors. In the case of the telescope, a large lens, or a large mirror, of small curvature collects as large an amount of light as possible from a distant object but magnifies the latter only slightly. In the microscope a small lens of great curvature produces high magnification of a small, near-by object but requires that the object be illuminated with a bright light. These instruments have been equally valuable. They rank among man's greatest tools of investigation.

The refraction of light by lenses is responsible for another optical instrument of great importance to science, commerce, and industry; the camera. The camera employs a condensing lens to focus the image of an object on a light-sensitized plate or film housed in the rear part of a lightproof box. It has a controllable shutter to allow light to enter the lens when desired. There is also a controllable iris diaphragm which will allow the correct amount of light from the object to pass through the

lens and expose the plate properly. The lens is fastened to a lightproof bellows so that it can be moved forward or backward in relation to the sensitive plate in order that the image of the object to be photographed may be sharply focused upon the plate. Although the optical system of modern cameras is very complicated, its fundamental purpose is merely to get a sharp image of the objects to be photographed on the plate or film. These objects are never all in the same plane, or the same distance from the camera; therefore, the camera must be designed to give a certain "depth" of focus. This is accomplished by varying the aperture of the lens by means of an iris diaphragm. The smaller the aperture the greater is the depth of focus and the sharper will be the details in the image of a scene or object with considerable depth. These requirements must be met in all cameras in which clearly defined pictures are to be made, whether they are the inexpensive kind for making "snapshots" or the most expensive ones for making the best motion pictures.

The Eye as an Optical Organ

Our eyes depend upon the refraction of light by lenses to give us sharp vision; and to us the eye is, of course, the most important of all optical devices. It is similar in many respects to a camera, in that it contains a condensing lens which is self-focusing, a lightproof box, an iris to control the amount of light entering, a shutter (the eyelid), and a light-sensitive area at the back. The lens is focused by increasing or decreasing its curvature, rather than by any movement forward or backward, as in the case of the camera, this focusing being produced by muscular action which changes the shape of the lens.

A sensitive surface, corresponding to the light-sensitive plate in the camera and known as the retina, is located at the back of the eyeball. Unlike the camera, however, the eye sees clearly at only one small spot directly in the line of sight, because the retina is so constructed that light falling on this small spot only produces clear vision. It is, therefore, impossible for a person to see clearly an entire page at any one instant, as a camera is able to photograph it in sharp detail with one exposure. This lack of uniform sensitivity of the retina is mitigated

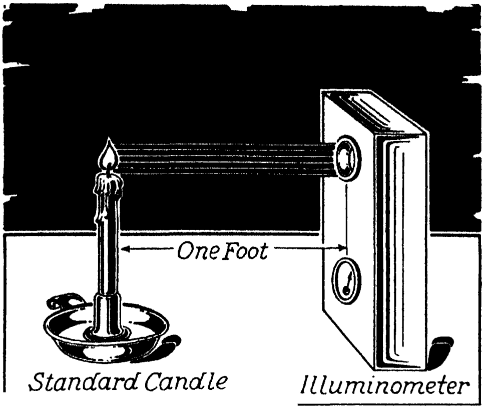
in human sight by the ability of the eye to turn easily and rapidly in its socket so as to scan the various details in a scene. Although the camera is able to make a picture of an entire page instantly, the eye sees the letters separately or, at best, a few words at a time; but by being able to move quickly, a series of successive images may be produced on the retina so that the whole page may be read.

When the lens of the eye is imperfect, eyeglasses containing corrective lenses may be used for aiding defective vision. Nearsighted people's eyes have natural lenses with too much curvature and therefore tend to focus the rays in front of the retina. Concave glasses worn in front of the eye can correct this, as the concave lens scatters the rays sufficiently to cause them to be focused farther back on the retina. Farsighted eyes have lenses with too little curvature so that the rays tend to be focused back of the retina. Convex lenses correct this defect by aiding in a shorter focusing of the rays. Eyes with natural lenses producing astigmatism need cylindrical lenses to correct their focusing. Sometimes combinations of cylindrical and concave or convex lenses are made to correct complex sight imperfections.

Practical Artificial Illumination

Since vision is affected only by reason of the light coming from an object producing an image in the eyes, no object can be seen without light. Likewise, too much or too little light may interfere with correct vision and produce eyestrain. The "science of seeing" has been extensively investigated in lighting engineering circles during recent years, especially with a view to determining the amount and the kind of light necessary for proper seeing conditions in the home, office, school, and factory.

In order to set any arbitrary measuring units for light, it is necessary to select a standard and reproducible source of light that will always have the same intensity. One of the first standards to be adopted was the "standard sperm candle," of a given size which would burn a specified weight of wax each hour. It has since become the "international candle," and it represents a source of light of one candle-power intensity. More permanent and conveniently operated electrical sources of standard candle power have since been developed. The intensity



One unit of measurement of light is the foot-candle, which is the intensity from a standard candle one foot away.

of light measured at a distance of one foot from the international candle, or its electric-lamp equivalent, is one foot-candle. The foot-candle constitutes, then, the practical unit of illumination. The instrument for measuring the intensity of light is the "illuminometer," or foot-candle meter. The modern form of this instrument contains a photoelectric cell which generates an electric current proportional to the intensity of the light falling upon it and an electric meter which measures this electric current. Instead of having a scale that reads in volts or amperes, however, the meter scale is calibrated to read directly in foot-candles.

A well-known fact in lighting engineering circles (although usually not recognized by the layman) is that different kinds of activities require different amounts of light for best vision and that for any activity involving the continued use of the eyes there is a minimum limit in the intensity of light below which severe eyestrain and probably permanent injury to the eyes will result. At least 10 to 20 foot-candles of artificial light are necessary for activities like casual reading. For critical reading, detailed benchwork, or fine needlework the amount necessary is about 50 foot-candles; and for work requiring that the eyes see very fine details, as much as 100 to 200 foot-candles should be used. As units for comparison, it might be noted that bright daylight is about 2,000 foot-candles; sun-

light, about 8,000 foot-candles; and moonlight, about 1/50 foot-candle.

The United States has been called a nation of people behind spectacles, and the statement contains a measure of truth. It is exceedingly fortunate that people who need glasses have such facilities available to them; yet the unfortunate part of the situation is that so many should require optical aid to give them satisfactory vision. Many accurately conducted studies have shown that one chief cause of defective vision is working under improper lighting conditions. This is particularly true of students and office workers who have long used too little light in study rooms and offices.

An important consideration in artificial illumination is that the intensity of light produced on a surface from a point source varies inversely as the square of the distance from the source to the surface. This relation between intensity and distance is conveniently expressed as

$$\text{Intensity in foot-candles} = \frac{\text{candle power of source}}{\text{distance in feet squared}}$$

In order to see how this relationship works out, consider the intensity of illumination on a surface five feet from a luminous source of 100 candle power. In this case we have 100 divided by five squared, or 25, giving four foot-candles. Should the luminous source be moved ten feet away, the intensity on the surface would be 100 divided by 100, or one foot-candle. In other words, the intensity of the illumination on a surface is reduced to one-fourth its former value when the distance from the source is doubled.

For good illumination it is also important to have a reasonably uniform amount of light shed over a fairly large area, so that no great contrast of light and dark is experienced in shifting the eyes from one spot to another. The eyes experience considerable strain in adjusting themselves quickly to looking from a brightly illuminated area to a poorly illuminated one. If the lighted area is particularly bright, glare is produced, which not only makes for bad seeing but also brings about eyestrain and possible eye injury.

To correct some of the inadequacies in artificial lighting, a new type of lamp was recently developed by the Illuminating

The I.E.S. study lamp with a 100-watt bulb produces ample light over a wide area without glare or high lights and deep shadows. (Science Service photograph.)

Engineering Society, usually referred to as the I.E.S. study lamp, for home and study-room use. It is designed to give indirect illumination of twenty to forty foot-candles, depending upon the size of the bulb used, over a fairly wide field. It has an indirect component that wipes out dark shadows beyond. Also, there is an absence of glare in the illumination from such a lamp that makes for ease in reading and helps to eliminate eyestrain.

Quality of the light used is another important item in good illumination. For ordinary vision, a light approximating sunlight is best; this, as we shall see later, is especially true when examining color. However, for sharpness of vision, for seeing details, a light as nearly as possible of a single wave length, called a monochromatic light, should be chosen; and this is

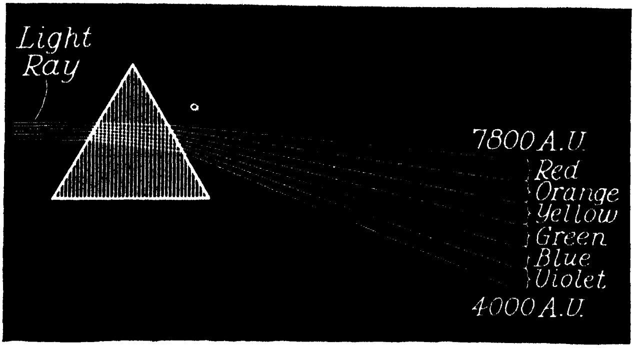
one of the reasons for the development, a few years ago, of the sodium-vapor lamp. It produces a light of strong yellow color and of a small band of wave lengths. It gives a sharply defined image in lenses, even in human eyes, owing to the lack of aberration, which is present when a large number of light waves of different wave lengths are used. Ordinary white light contains many wave lengths and because of color aberration produces a slightly blurred or imperfect image when passed through ordinary lenses; the eye lens is no exception to this effect. Aberration occurs because each color, after going through the lens, is focused at a slightly different position because of a slightly different index of refraction for different wave lengths. Light of a pure color eliminates this trouble because it contains only a relatively few wave lengths, all of which are focused at the same point.

The first practical use of sodium-vapor lamps in the United States was for highway lighting in an installation along the Balltown Road between Schenectady and Albany, N. Y., in 1933. About one year later, a long installation was made on Central Avenue in New York City. Sodium-vapor lamps provide the clearness of vision and sharpness of detail necessary for greater safety in driving. They have now been adopted for highway lighting along many roads and streets of the nation, and they have been found particularly useful in tunnels, on long bridges, at busy intersections, and at railroad crossings.

A lamp producing a predominance of blue rays is the Cooper-Hewitt mercury-vapor lamp. It is used in metalworking factories and machine shops, where it shows up great detail in intricate machined parts. In the future it may not be uncommon to see pure yellow light used in offices and libraries, pure blue light for metalworkers and watchmakers, and other pure colored lights in other industries.

The Nature of Color

Probably the most common and esthetically significant characteristic of light is its ability to produce the sensation of color. Color in nature is so widespread and universal that it is difficult to imagine a world of only white, black, and gray. When we look at an ordinary photograph we see this neutral



White light may be separated into six colors.

sort of world; and if we could eliminate all memories of the hues of nature that we associate with the objects in the picture, we should become fully conscious of the drabness that would result from living in a really colorless world. The dulling of nature's colors on a cloudy or rainy day and their gradual and complete fading as night falls is direct evidence that color is a characteristic of light. Since the time when Sir Isaac Newton in 1666 caused a beam of sunlight to pass through a glass prism and produced a continuous gradation of colors of the spectrum, light has been known for this property.

When a beam of white light is passed through a glass prism, it is separated into a spectrum of six major colors, red, orange, yellow, green, blue, and violet. The classification is an arbitrary one, as actually each major color blends into those adjacent to it. In some classifications a seventh major color, indigo, is recognized as falling between blue and violet, and many other intermediate hues are distinguishable. When the wave lengths of the six major colors are measured, they are found to range from about 7,800 angstroms for the red to about 4,000 for the violet, as represented in the accompanying drawing. In fact there is a continuous range of several thousand wave lengths from the longest to the shortest of those mentioned above; the eye, however, is unable to distinguish this large number as different hues. Apparently the eye is able to distinguish only a few hundred different hues, and these include an exceedingly small number of pure colors and a larger number of combinations that merge almost imperceptibly into each other.

White light is a mixture, therefore, in reasonably equal proportions of all wave lengths of the visible spectrum; and the fundamental cause of color is the difference in wave length of the various light waves. The most familiar and splendid example in nature of the breaking up of white light into the colors that compose it is the rainbow. It is to be observed in the eastern sky in the late afternoon during a rainstorm and when the sun is shining through the clouds, and it may also appear in the western sky in the morning under the same conditions. It is seen in a waterfall, in the spray of a garden hose, or whenever sunlight plays at the proper angle on falling water drops. In each case the display of colors is produced by the droplets of water separating the sunlight into its different wave lengths as it passes through them. The same result may be obtained when sunlight or any other white light is passed through a prism. The separation is produced by an unequal refraction of the different wave lengths in which the shorter waves are bent more than the longer ones in passing through the medium. This is color from light, and in the discussion to follow the reader should bear in mind that light color is distinguished from pigmental color in its primaries, its complementaries, and the manner in which it must be handled to produce a given effect.

Should the colors secured by separating white light into its spectrum be recombined and allowed to fall on a screen, white light would again be produced. This process of securing white light by combining the spectrum colors in proper proportions was first discovered by Newton and has been easily verified since his time. Furthermore, it has been found that the sensation of white light is produced by the addition of red, green, and blue-violet lights; and it was long ago demonstrated that any color could be secured by the proper mixing of two or three of these colors. The colored lights of red, green, and blue-violet have been designated the primary colors. Mixing the colored lights to produce other colors or white light is called the "additive" process of color mixing. By observing the color chart representing the primary colors of light, shown facing page 370, it is seen that green added to blue-violet gives a blue-green, that blue-violet added to red produces a magenta, and that red

combined with green results in yellow. Since yellow may result from a combination of red and green, it is possible to add yellow and blue-violet lights to produce white light.

Another way of describing the additive process of color mixing is to note that if red is removed from white light, a blue-green is secured; when green is removed, magenta results; and when blue-violet is removed, yellow is secured. The three colors produced by removing separately the three primary colors are referred to as the complementary colors. A complementary color, therefore, is that color sensation which we perceive when one color is extinguished from white light. For example, when the proper filter for absorbing blue-violet is inserted into a beam of white light, the light passing through is yellow; yellow is said to be complementary to blue-violet. The definition of a complementary color would indicate that many different pairs of colors are complementary to each other. The three particular colors that are complementary to the three primary ones are sometimes referred to as the secondary colors. The relationship of the complementary, or secondary, to the primary colors of light is conveniently expressed as follows:

Primary Color	Complementary Color
Red	Blue-green
Green	Magenta
Blue-violet	Yellow

By using combinations of the three primary rays of different rather than equal intensities all the various intermediate hues of the rainbow may be produced. The additive process of mixing colored rays, however, is not the method by which most objects and substances of nature produce color. How, then, are the colors in objects, fabrics, paints, and dyes produced? The answer is a definite and specific one but not difficult to understand.

Most of the objects that we observe are visible because of "borrowed" light which they receive and reflect. When the source of the reflected light is extinguished, the object disappears from view. Since sunlight is white, it is easy to understand that any object that reflects all wave lengths equally will itself appear white or a very light gray and that any object that absorbs most of the rays in equal proportion will appear

black or a very dark gray. Most substances in nature, however, do not have the property of reflecting or transmitting all wave lengths of light to the same degree; rather, they absorb some wave lengths and reflect or transmit others. When white light falls on a surface that absorbs some wave lengths and reflects others, the object will be seen by the light that is not absorbed and will have the color of those waves which remain. Thus, color in objects is due to a selective absorption (or subtraction) of certain waves from white light. In other words, the colors that we see in objects are, in most cases, composed of those rays which are not subtracted from white light when the light falls on the object. This method of obtaining color is, therefore, a "subtractive" process. For example, a red object appears red because it contains a pigment that has the property of subtracting from white light all wave lengths except those of red, which it reflects or transmits to the eye. Likewise, an object is blue because its pigment absorbs all colors except the blue, which is reflected or transmitted; and similar conditions hold true for other colored pigments.

Pigmental Colors

Most of the color observed in nature and the color produced in painting and printing is what might be called pigmental color; *i.e.*, it is color produced by the selective absorption of various pigments. The mixing of various pigments to produce different colors is a very different procedure, therefore, from that of mixing colored lights to produce these colors. When yellow and blue lights are combined, white is secured; but every schoolboy knows that mixing yellow and blue pigments will produce not white but green. This may readily be understood when it is remembered that in the addition of color lights, the tendency is to build up white, since the actual energies of the different wave lengths are being combined; however, combining pigments tends to produce black, since each pigment subtracts some of the color of the received light. Yellow pigment subtracts from white light all colors except yellow, some orange, some red, and some green. Blue pigment subtracts all colors except blue, some violet, and some green. When these two pigments are mixed in proper proportions, the only light not

absorbed by the combination is green; and since this is reflected or transmitted, the result obtained is green.

It becomes clear, therefore, that mixing light of given colors produces an entirely different result from the mixing of pigments of corresponding hues. Colors mutually complementary as light may not be mutually complementary as pigments. When two colors that as lights will give white are mixed as pigments, they will not produce a light gray, which is the nearest approach to white that is possible from the mixture of colored pigments; rather, they are likely to give another color. The artist uses as his primary colors the three that are not producible by mixtures of other pigments and that themselves will give the greatest range of hues when they are combined. These pigmental primaries are crimson red, yellow, and blue. The secondary hues (*i.e.*, those produced by a mixture of equal parts of two primaries) are orange, green, and violet. The relationship of the pigmental primaries and secondaries is conveniently expressed as follows:

Primaries:	Crimson-red Yellow	Yellow Blue	Blue Crimson-red
Secondary:	Orange	Green	Violet

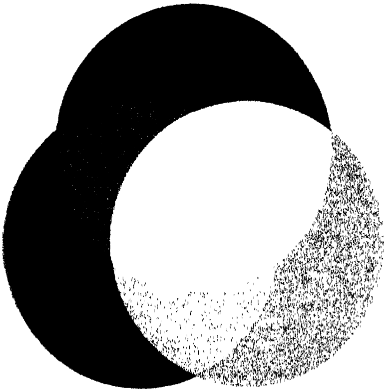
When all three of the pigmental primaries are combined, the result is a dark gray, the nearest approach to black that it is possible to secure with pigments. On the other hand, should unequal amounts of either the primary or the secondary pigmental colors be mixed, all the intermediate hues observed in nature are obtained; but in each case it is necessary to remember that the color obtained will be what is left from the reflected or transmitted white light after all the colors of the mixed pigments have been subtracted. In the accompanying color chart the pigmental primaries as well as the secondaries produced by their mixing are represented in the diagram at the upper right. Just how black is the center where the three primaries overlap will be determined by a number of factors, not the least important of which is the quality of inks used. The diagram at the upper left represents the light primaries and the various hues secured by adding the primaries. In the additive process it is seen that the three primaries combine to give white light.

From a theoretical viewpoint, a black object would be one that absorbs all the light falling on it. No substance, however, is such a perfect absorber. The best blacks reflect about 4 per cent, and also black pigments are never wholly free from a small amount of selective absorption which gives them a tinge of brownish or bluish color. Likewise, a pure white object should reflect 100 per cent of the light falling on it. But, again, the perfect is unobtainable, and a good white is one that reflects about 90 per cent of the incident light.

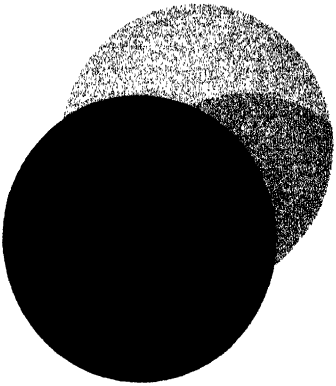
Perhaps we may understand better some of the qualities that we see in the colors of nature if we note that most objects reflect in two ways the light that falls upon them. First, some of the light is reflected from the surface without change; when this happens, white light will be reflected as white regardless of the color of the object. This type of reflection is called specular reflection; and there are various degrees of it, from that of a perfect mirror in which all light is so reflected, to that of a mere sheen of white light which mingles with the color of the object, as on a polished opal. In fact some light is reflected specularly from practically the whole of any illuminated surface. For example, because of this surface reflection it is almost impossible to see the colors of an oil painting from certain positions in a room. Second, any light that is not specularly reflected passes to a greater or lesser distance beneath the surface. Selective absorption takes place while this light is passing through the molecules of the pigment; and when it emerges, either by reflection or by transmission through the object, it is deficient in all wave lengths except those we see as the color of the object.

Thus, the total color quality of an object is the result of its specular reflection of white light combined with its reflection or transmission of colored light after the selective absorption produced by the pigment takes place.

Objects that show their colors by transmitted light are usually brighter and have purer colors than objects that produce their colors mostly by reflection. For example, a piece of red glass when viewed against a dark background and with the light falling upon it from the same side as the person who is viewing it, so as to be seen entirely by reflection, may look



Light primaries and the colors produced by their combinations



Pigmental primaries and the colors produced by their mixing

These colors differ chiefly in hue

These colors differ chiefly in value

These colors differ chiefly in chroma

dull or somewhat gray. Yet when the glass is held between the light and the eye, so that the rays are transmitted, the red is much brighter and usually more brilliant. This results from two causes. In the first place, the transmitted rays pass directly through the red glass from the source of light, whereas the reflected rays are somewhat diffused and also loose some of their energy by reflection. In the second place, the specular reflection at the surface takes place only on the side toward the light. The white light so reflected is mingled with the reflected color and tends to dilute the red color of the glass, producing the grayish effect; however, in viewing the glass by transmitted light, no specular reflection is seen, as it is entirely on the opposite side, and the color is much purer. These differences are particularly noticeable when a stained-glass window of a church or the color in autumn leaves is seen by transmitted light and then by reflected light. When the stained-glass window is looked at from inside the building by transmitted daylight, the illumination is much greater and the hues are much more brilliant than when the same window is seen from the outside entirely by reflected sunlight. The colors in the leaves on a tree take on a brilliance and purity of hue when the sun is seen shining through them that far surpass the same colors seen only by reflected sunlight.

Since most of the color that we see in nature is produced by the selective absorption of certain wave lengths and the reflection or transmission of others, the quality of light falling upon an object has a marked effect upon the color that is seen. The natural color of objects is considered to be that which they display when viewed under daylight, or light that contains all wave lengths of the spectrum. When multicolored objects are looked at under light that is deficient in some wave lengths, some of the colors will be missing, and an entirely different effect will be noticed. Anyone who has purchased delicately shaded articles of a variety of colors in a store illuminated entirely by artificial light is likely to discover a new and sometimes quite noticeable color pattern in such articles when they are seen by daylight. Artificial light is usually deficient in the blue and violet colors of the spectrum and contains an excess of

red, orange, and yellow. Colored objects observed under such light will have their blues subdued and deadened, but the reds and yellows will be particularly noticeable through overemphasis.

The explanation of this effect is that before a color can be reflected or transmitted to the eyes it must be present in the light falling on an object. A red object will look red when seen under white or red light. In each case the red color is present in the light, and it is reflected or transmitted by the pigment of the red object. Such an object looks black, however, when looked at under pure green, blue, or violet lights, for these contain no red color to be reflected or transmitted, and the pigment absorbs the green, blue, and violet colors. Under such conditions no light is transmitted to the eyes except a small amount by specular reflection, and the object looks black. Likewise, a blue object looks blue under white and under blue light and looks black under red light. In this case the blue colors of white or blue lights are reflected or transmitted by the object so that they may be seen; but the red color of the red light is absorbed; and since the red light contains no blue color, no light is reflected, and the object appears black.

Characteristics of Color

Most people are able to recognize differences in the colors of nature, whether these be the colored rays of light in the rainbow or the pigmental colors of objects resulting from selective subtraction of certain wave lengths from white light. However, an attempt on the part of the average layman to describe these differences as manifested in the beautiful colors seen in the sky at sunset, in flowers, in a painting, or in a tapestry is usually vague and often ineffective. It is common practice to use such terms as "red as a rose," "sky blue," "emerald green," as well as a host of other phrases, in attempting to convey an idea of color. Obviously such terms are subject to various interpretations. But a few characteristics of color will, if understood, give a better knowledge of color differences and a more specific language with which to describe them. These apply to colored light rays, and they have their analogous characteristics as applied to color produced from pigments. Characteristics of color as light are hue, brightness, and satura-

tion; and the analogous characteristics of pigmental colors are hue, value, and chroma.

Hue as applied to colored rays is suggested in the name of the color, such as red, green, or violet. Hues are usually represented by a wave length of the spectrum; for example, yellow light from the sodium-vapor lamp has a wave length of about 5,890 angstrom units. Some hues are represented by composite spectral colors, purple being a prominent example. There is no single wave length in the spectrum that will produce a purple hue; but when the proper wave lengths of red and violet are combined, purple light is secured.

Let us see what is meant by the brightness of color as light. When a white light shines on a white screen, the brightness of the screen depends upon the light. When it is increased to its greatest possible amount, the screen reaches the brightest white; and as the light is slowly dimmed to zero, a series of neutral grays are noticed, which range from the brightest white to the darkest black. This series of neutral grays is used as a standard for judging the brightness of a light of any hue or color; *i.e.*, the brightness of a given color is said to be that of the neutral gray that appears to have the same intensity.

Saturation of a color is determined by the amount of white light that is mixed with the colored light; and complete saturation exists when no white light is present. Saturation is, in a sense, the measure of the color's purity, and the greatest saturation obtains when the wave lengths of the specific color only are present. When white light of the same brightness is added to a color, the brightness does not change, but the color becomes faded or washed out; and should sufficient white light of the same brightness be added, the saturation would become zero, and the color disappear.

The artist's special problem is pigment mixing with its resultant effects, however, and not light mixing; also, most of the colors that we observe in nature are pigmental colors. Pigmental colors have certain characteristics that are analogous to, but not identical with, the characteristics of color as light. A possible exception involves hue. The artist uses the term hue to specify any chromatic color, *e.g.*, the three pigmental primary hues of crimson-red, yellow, and blue, or the secondary hues of

orange, green, and violet. The hue of a pigment may be changed by mixing with it a pigment of another color. A large number of intermediate hues may be secured in this manner, several hundred of them being distinguishable by persons who are not color blind.

The other characteristics of pigmental color are value and chroma. Value in any color refers to the relation of the color to white and to black, which we indicate when we say a "light blue" or a "dark blue." By mixing white or black pigment with a color, the value of the color is changed without changing its hue. By changing the value of a given color, it is possible to secure a large number of different tints and shades of that same hue, and these may be extended by the addition of white or black to the limits of the eye's ability to distinguish one tint from its adjacent one.

One other characteristic of pigmental color is not described by either its hue or its value. This is what the painter and the printer refer to as chroma. Perhaps the simplest way to define chroma of a color, although the words are scarcely adequate, is to say that it means the color strength of a hue as compared with a neutral gray. It is used to distinguish strong from weak colors of a given hue and is implied when we use such expressions as "brilliant red" or "dull red." Chroma means the purity of a given color, *i.e.*, its freedom from a neutral gray. When it is desired to weaken a color, to "kill some of its fire," without making it lighter or darker and without changing its hue it is necessary to add both white and black pigments in the proper amounts. Thereby the chroma is changed, but the value and the hue remain the same. A color close to gray is thought of as being a weak chroma; and as the color gets farther away from gray toward the purest color of the same value, its chroma becomes stronger.

The illustrations at the bottom of the color chart facing page 370 are included in order to help visualize the meaning of hue, value, and chroma. The three colors at the top differ in hue; the three center ones are of the same hue, red, but differ in value from a light to a dark red; the three lower colors differ chiefly in chroma or purity, the chroma becoming weakened by the addition of neutral gray.

With hue, value, and chroma accurately defined by color experts, it is possible for them to represent different degrees of these characteristics on a numerical scale and thereby obtain a specific and definite language for describing color. One of the most widely used systems of color notation employs ten principal divisions of hues (represented by letters), ten numbers for value (representing black with zero and the greatest white by ten), and fourteen steps for chroma (with neutral gray being zero and the most brilliant red being fourteen). By means of such a system, any specific color may be easily identified in terms of hue, value, and chroma. For example, a brilliant red and one of the strongest available pigments is R 4 /14, in which R shows the hue, 4 represents the value, and 14 shows the chroma. The designation G 8 /3 shows that the color is green, of value 8 (or a light pastel), and the chroma 3, which puts it close to gray. This system (the Munsell) is mentioned here not with any thought that the reader will master it to describe the colors of nature but rather to point out that color experts may describe and use color quantitatively by employing this or similar systems.

Printing in Color

Some of the color reproductions seen in magazines and books are objects of beauty and works of art as well as examples of scientific and technical achievement. The finished pictures give little indication of the steps necessary to produce them; only the skill is evident. The art of color printing is a very accurate and technical process that is now well understood. It may be said to be one of the most highly developed phases of the printing technique. As such, color printing is a foremost consideration in advertising and in the reproduction of famous paintings; its use is becoming widespread in magazine illustrations; and it is beginning to creep into book illustrations.

Finest of all the usual color reproductions are those produced by the "four-color process," in which three different colors and black are applied in their proper places on the paper by four separate steps of printing. Each step requires the use of a separate halftone plate for applying the desired colored ink. The process of color printing involves, first, the securing of

the halftone plates to represent the primary hues in the scene to be reproduced, and, second, printing these hues in proper register so that the eyes see not only the primary colors in their respective places but also intermediate hues by a "visual" mixing of the color from the dots carrying the separate primaries. The white in the scene is furnished by the paper, of course, and the black halftone is used to provide the grays as well as to add greater contrast to the details of the picture.

The separate plates of the four-color set are made by photographing the scene or painting four times, each time through a translucent filter which transmits one primary color or the grays. From each of these negatives a metal plate is photoprinted and etched in such manner that the areas for a given color consist of raised dots on which is applied the ink that is to be transferred to the paper in printing. The number of dots in a given area will be in proportion to the amount of the color present in the copy at that place. In arranging the patterns of the dots on the separate plates an attempt is made to have them lie side by side rather than superimposed on each other when the four plates are printed. This result is accomplished to a remarkable degree by rotating the screen angle at the time the dots are made on the plate; however, there is some overlapping because of the large number of dots per square inch in a good four-color reproduction. The production of intermediate hues is thereby a visual one rather than an actual blending of the inks. All of these facts may be noted by looking at any four-color reproduction through a magnifying glass.

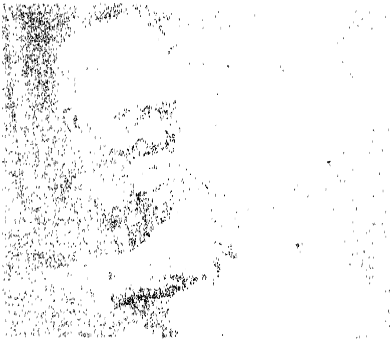
In building up the finished picture the yellow ink is applied first, then the red, next the blue, and finally the black. These steps are strikingly shown in the action photographs on the facing page in which the four separate color plates are printed at the left, while at the right the progressive addition of each color is shown from top to bottom with the last one being the finished picture.

REFERENCES FOR MORE EXTENDED READING

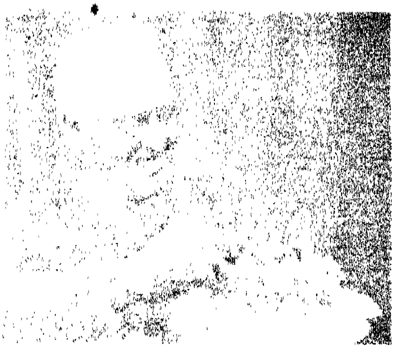
LUCKIESH, MATTHEW: "Torch of Civilization," G. P. Putnam's Sons, New York, 1940.

Written by an authority, this book is a fascinating account of what lighting has done for man.

Yellow



Yellow



Red

Yellow and Red

Blue

Red, Yellow and Blue

Black

Red, Yellow, Blue and Black

Photographs showing progressive steps in four-color printing process

EDGERTON, H. E., and JAMES R. KILLIAN, JR.: "Flash!" Hale, Cushman & Flint, Inc., Boston, 1939.

Some of Dr. Edgerton's extraordinary photographs taken by ultra-high-speed photography are here reproduced with interesting explanations of how they were taken. Not only a valuable book on the techniques of photography but also, because of the nature of the objects photographed, an excellent one for scientific cultural knowledge.

MINNAERT, M.: "Light and Color in the Open Air," George Bell & Sons, Ltd., London, 1940.

Simple as well as complicated observations are described in language of the average person's understanding of such things as sunlight and shadows, reflection and refraction of light, the curvature of light rays in the atmosphere, colors, rainbows, haloes, luminous plants, animals, and stones.

EYRING, C. F.: "A Survey Course in Physics," Prentice-Hall, Inc., New York, 1936, Chaps. XII–XIV, inclusive.

The beginning student will find these chapters an excellent treatment of the fundamentals of light, color, and optical instruments.

MACK, J. E., and M. J. MARTIN: "The Photographic Process," McGraw-Hill Book Company, Inc., New York, 1939.

The first four chapters contain an excellent discussion on image formation by various types of lenses as well as interesting material on the history of photography and modern cameras.

"The Photography of Colored Objects," Eastman Kodak Co., Rochester, N. Y., 1938.

An excellent little book on the optical properties of colors and the principles to be considered in photographing objects in color.

SOUTHALL, J. P. C.: "Mirrors, Prisms and Lenses," The Macmillan Company, New York, 1933.

Although this is a college textbook of geometrical optics, it is an excellent reference for the intelligent layman who wishes some specific information on the optical properties and uses of lenses, mirrors, and prisms.

SARGENT, WALTER: "The Enjoyment and Use of Color," Charles Scribner's Sons, New York, 1928.

This standard and popular text for art students forms a valuable reference for the layman interested in the use of color in its many settings.

"Monographs on Color," International Printing Ink Corporation, New York, 1935.

The printing industry has produced three small volumes on the chemistry, physics, and use of color, beautifully written and extensively illustrated in color.

BLUM, H. F., "Photodynamic Action and Diseases Caused by Light," Reinhold Publishing Corporation, New York, 1940.

The first chapters of this very illuminating book contain an excellent introduction to the elementary principles, including the quantum aspects, of radiation. The reader will probably continue through the latter part, dealing with diseases caused by light.

HARDY, A. C., and F. H. PERRIN: "The Principles of Optics," McGraw-Hill Book Company, Inc., New York, 1932.

An advanced text for college students who are training for a career in physics or optical engineering; it presents such a broad and authoritative treatment of the field of optics, however, as to make it a valuable reference for others who wish specific information regarding optical phenomena or optical instruments.

Journal of the Optical Society of America, published by The American Institute of Physics, New York.

Technical and semitechnical articles on a wide variety of topics relating to light and optics.

Illuminating Engineering, published by Illuminating Engineering Society, New York.

Those interested in the technical development and professional news of the lighting industry will find this journal a valuable reference.

Westinghouse.

12: INVISIBLE RADIATION

Or the Five Other Divisions of the Electromagnetic Spectrum

THE electromagnetic spectrum consists of a series of five invisible divisions in addition to visible light. One of them furnishes the earth with heat and is the source of much of our mechanical power. One helps living creatures to develop strong and normal bones. One gives us insight into the structure of material things. One helps us to fight diseases and has aided in discovering some of the properties of matter. One helps man to conquer space and to fling his messages around the earth.

Such potent energies as these have been labeled with specific, and sometimes descriptive, names, although in certain cases titles have been gradually adopted or relate to some incident in their discovery. In the order referred to above, the spectrum

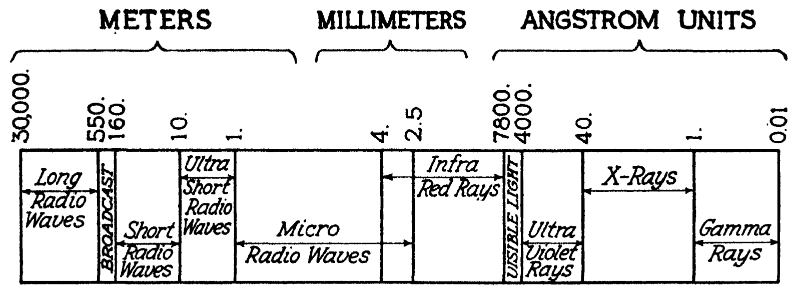


Chart to represent the main divisions of the electromagnetic spectrum.

divisions of invisible radiant energy are infrared waves, ultraviolet waves, X rays, gamma waves, and radio waves. These five, together with visible light, constitute the entire range of electromagnetic radiations known to man.

Borderland to Light

Ultraviolet radiation might be introduced as the "health-giving" band of the electromagnetic spectrum because it aids in producing one organic substance necessary to the growth of animal creatures. It is emitted by the sun along with vast quantities of visible light and heat waves. Ultraviolet rays range in wave length from about 4,000 to 40 angstrom units. They are therefore shorter in wave length than the violet rays of the visible spectrum. Their place in the electromagnetic spectrum is immediately adjacent to the short-wave-length end of visible light.

Ultraviolet radiation may be produced artificially in various types of lamps, the best known and most widely used of which is the mercury arc. This is filled with mercury vapor and so constructed that the gas is activated by a stream of electrons passing through it from a hot filament, or in an electric arc. In the process a kind of activation occurs in which the electrons in the outer shells of the mercury atoms are caused to jump to different energy levels within the atoms by the bombarding effect of the swiftly moving electrons from the filament or in the arc. Their movement between the different energy levels of the atoms causes the atoms of mercury vapor to emit ultraviolet radiation. A carbon arc also will produce ultraviolet radiation along with large amounts of visible light and radiant

heat. In this lamp, too, the carbon atoms are so activated by the passage of the electric current that ultraviolet radiation is emitted. Ordinary tungsten-filament lamps emit little ultraviolet radiation; the small amount emitted by the filament, usually of wave lengths ranging between 4,000 and 3,000 angstroms, is absorbed by the ordinary glass envelope and is not radiated from the lamp.

A statement that ultraviolet radiations are easily absorbed by matter may seem abrupt and uninteresting; nevertheless, the fact is that large amounts of these rays emitted by the sun never reach the surface of the earth, as they are caught and absorbed by the upper atmosphere. This is a fortunate condition for us, for such intense radiation would surely soon destroy all life on earth. Ordinary glass absorbs most of the ultraviolet rays that do reach the earth's surface, and sunlight passing through glass windows no longer possesses the rays that produce sun tan and provide certain health advantages. Likewise, these rays are easily absorbed by most kinds of clothing; they will not pass far into water or other liquids; and they are absorbed by the outer layers of the skin.

Some materials, particularly special varieties of glass, transmit ultraviolet rays. Quartz transmits waves throughout most of the ultraviolet spectrum with the highest percentages of any of the glasslike materials. Vitaglass and Corex will pass the longer ultraviolet waves as well as some of the shorter wave lengths. It is necessary to use either quartz or one of the special glasses for the construction of solaria for ultraviolet treatment in order to permit the desired radiation from the sun to reach the patient. Optical instruments, such as cameras or spectrographs, employed for studying ultraviolet must use special lenses and prisms, preferably made of quartz, to allow the waves to be transmitted. Another glass, known as Wood's glass, will pass one band of ultraviolet radiation, although it will not pass visible light to any extent. An ultraviolet lamp surrounded by this glass will produce the so-called "black light," which means that visible light is absorbed and a portion of the invisible ultraviolet transmitted.

As indicated by reference to the camera and spectrograph, ultraviolet waves can be focused, and they obey the same laws

of refraction and reflection that are common to other waves. In fact they behave much as light waves do in all their characteristics except the manner in which they are absorbed. When the waves from a lamp that emits both visible and ultraviolet radiation are passed through a spectroscope having quartz lenses and prisms, the ultraviolet spectrum is produced adjacent to the visible spectrum. Should these spectra be focused on a screen containing a chemical substance that will fluoresce or shine with visible light where ultraviolet waves fall upon it, the ultraviolet spectrum will be beautifully shown. Ultraviolet rays may be used in photography; and when it is necessary to obtain exceedingly fine detail in a picture, ultraviolet rays rather than visible light are employed to "illuminate" the object. This means, of course, that ultraviolet rays will affect a photographic plate in the same manner as do light waves. The only difference between the two bands of radiation in this respect is that the ultraviolet waves are more actinic than the light rays. It is possible, therefore, to detect ultraviolet waves and to measure the intensity of the radiation by the fluorescent effect or by the photographic effect.

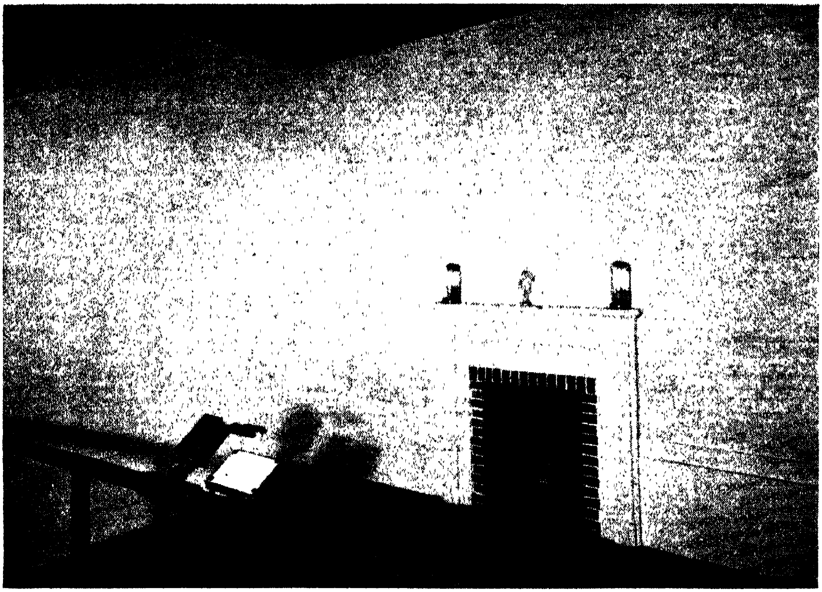
Ultraviolet rays produce in the fatty tissues of living creatures an important food element known as vitamin D. Just how this is accomplished is not definitely known, but exposure to ultraviolet radiation, either from the sun's rays or from special lamps, brings about the formation of vitamin D in the body. Rickets in children may be prevented or cured by sufficient vitamin D in the body system, secured either by exposure to ultraviolet radiation or by being consumed in foods or medicines that contain this vitamin. Proper exposure of the body to this radiation also seems to result in a general increase of resistance to germ infection. For example, there is some evidence to show that proper exposure will prevent, or lessen the severity of, common colds. The particular part of the ultraviolet spectrum that produces the formation of vitamin D and the added resistance to disease infection is the band from about 3,000 to about 2,900 angstroms, and this band is often referred to as the "vital" rays.

It is advisable, however, to warn that overexposure to ultraviolet radiation endangers health. This is true whether the

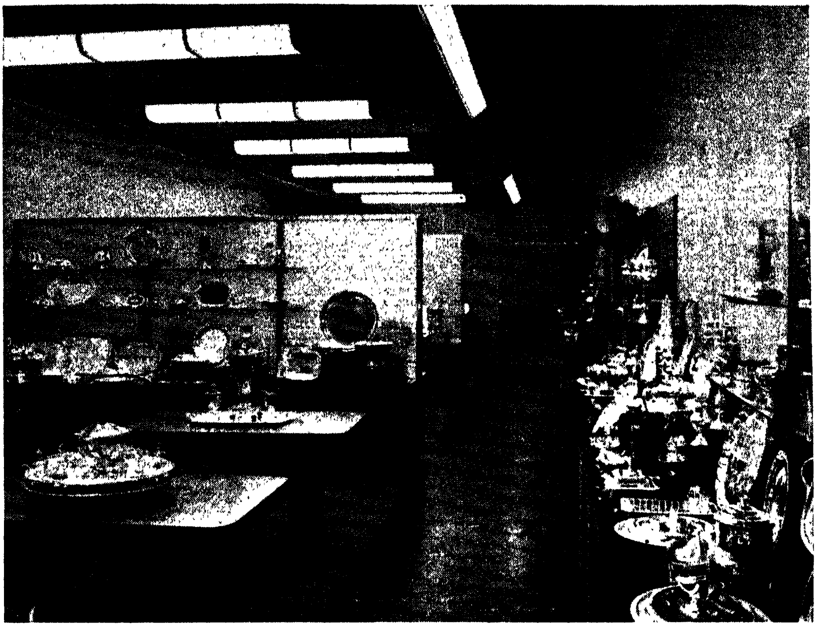
A striking illustration of the effect of ultraviolet radiation on bacteria. The right side of this culture was radiated and the bacteria destroyed, while rapid growth continued on the left side. (Courtesy of General Electric.)

exposure is from the so-called "health lamps" or from the summer sun's rays. Too great irradiation by ultraviolet rays produces painful sunburn, and the living cells of the skin may be destroyed. Severe sunburn may do much bodily harm, even to the point of causing death. A general weakening of the body, which affects the kidneys, is one danger; and damaging the eyes, which are very sensitive to this radiation, is another. In fact sun tan is a protective measure taken by the body against overexposure to the rays. Sun tan consists of pigmented cells formed in the outer layers of the skin which absorb the ultraviolet rays and thereby prevent them from passing deeper into the skin. A practical rule is never to use powerful ultraviolet lamps except on the advice and under the direction of a skilled physician and never to burn the skin by trying to get a deep tan in a few days.

Ultraviolet radiation has many industrial uses, some based upon the fluorescent effect. A quick identification of ores in mining and of drugs, foods, and oils can be made by exposing them to ultraviolet and noticing the colors of the fluorescent light. Certain ordinarily invisible inks become visible under this radiation, a phenomenon useful in criminology and banking. Another use of ultraviolet radiation is the practice of exposing certain foods to its action in order to create in them vitamin D.



Blank walls greet the visitor to this room in the Franklin Institute when looked at under visible light. When the lights are extinguished and the ultraviolet lamps turned on by operating a switch, luminescent murals appear from fluorescent pigments used to paint them. (Photographs by Gladys Müller, Franklin Institute.)



High-level illumination in a silverware showroom produced by white-light fluorescent lamps. (Courtesy of General Electric.)

Even tobacco has been processed by ultraviolet radiation to break down some of the acids and oils in it and to make it "milder."

One of the newest and perhaps most significant practical uses of ultraviolet radiation is in the new fluorescent lamps, in which mercury vapor is activated by an electrical discharge through it so that ultraviolet radiation is produced. The glass tube surrounding the lamp is coated on the inside with a material that will fluoresce with a visible color when the ultraviolet strikes its surface. By using different fluorescent coating materials any color or combination of colors may be produced. Fluorescent lamps offer not only a wider range of colors but also a whiter light or a closer approximation of sunlight than any other lamp known.

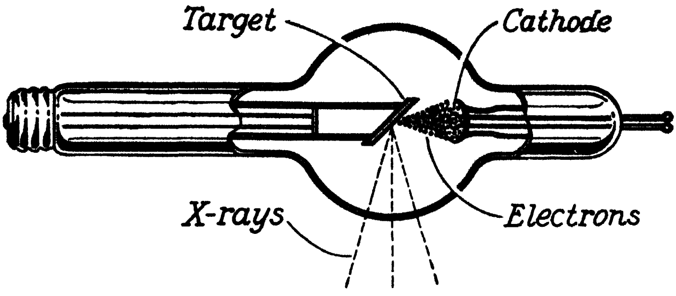
The fluorescent lamps lend themselves to a wide variety of uses because of the range of colors that may be produced. An additional factor in their favor is that their operating cost is approximately one-third to one-fifth that of the incandescent

lamp. They produce a relatively high brightness, yet there is a softness to the light and almost complete absence of glare. A large portion of the lighting of the New York World's Fair of 1939 and 1940 was produced by such lamps, which gave to the buildings and grounds a very unusual and pleasing effect at night. Installation of fluorescent lamps for home and office lighting was materially begun in 1940, and some divisions of the lighting industry predict that such lamps will become the most universally used type of artificial illumination of the future.

X Rays

Since their discovery by Roentgen in 1895, X rays have been found to be a part of the electromagnetic spectrum and to be adjacent to the ultraviolet in their wave lengths. In fact we now know that there is no sharp boundary line between these two divisions of the spectrum but rather that one shades into the other. The wave lengths of the X rays have been measured and found to extend from about 40 to about one angstrom unit. This makes their wave lengths of about the order of one-thousandth as long as those of ordinary light. Such small sizes enable them to pass more or less easily between the atoms of matter, thus accounting for the characteristic that led to their discovery and the one that has been most widely used, namely, their ability to penetrate dense substances.

Almost everyone is familiar with the fact that X rays are produced by an "X-ray machine." This device consists fundamentally of a special type of vacuum tube and the necessary apparatus for supplying the proper voltages to the electrodes of the tube. In it are two primary elements essential to its operation. One is a filament that can be heated by a small electric current, not greatly unlike the filament in an ordinary electric-light bulb. When it is heated, it gives off a stream of electrons, generally referred to as a cathode ray. The other element is a strip of metal, usually an alloy of tungsten or platinum, mounted on a copper holder which is referred to as the target. The target is situated directly in front of the filament and is connected by a wire to the outside of the tube so that positive voltages ranging from about ten thousand to a million volts may be applied.



Illustrating operation of an X-ray tube.

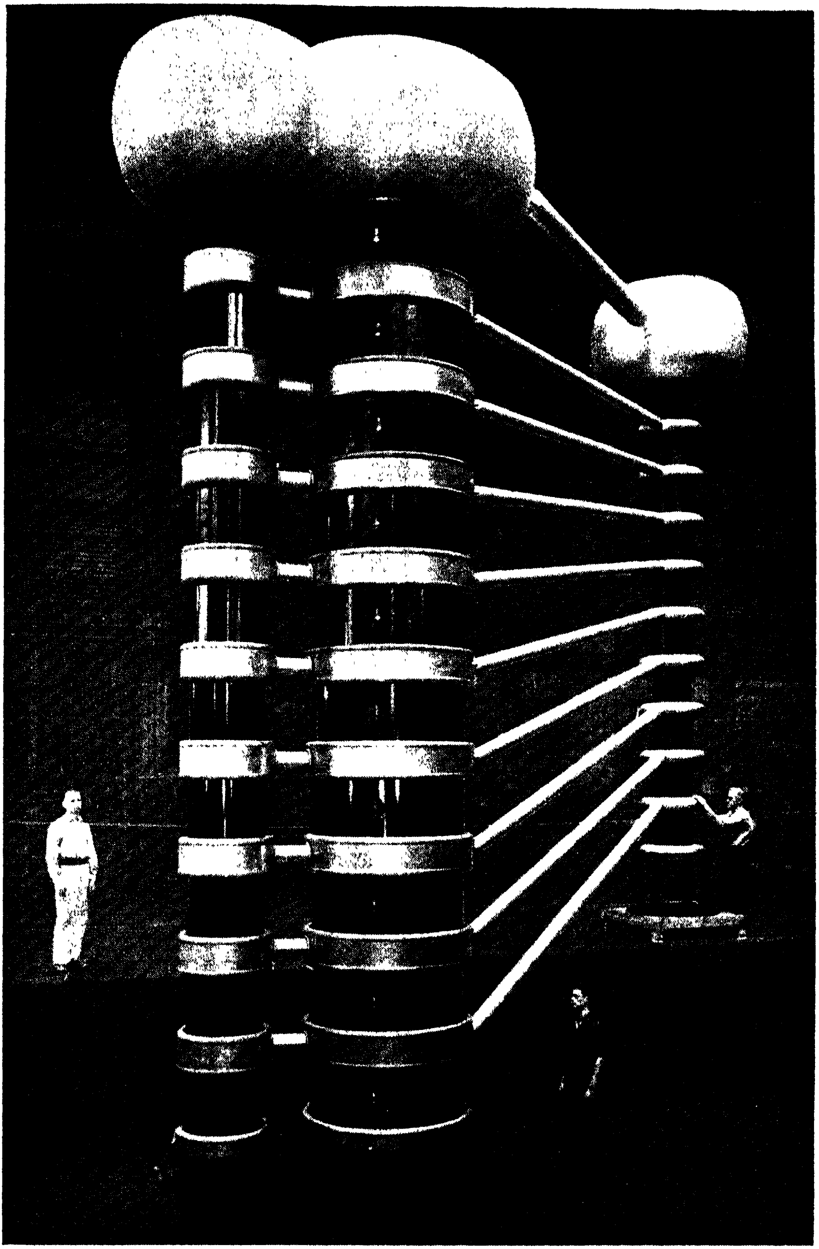
When the filament is heated and a high voltage is applied to the target, the electrons released at the filament are drawn across to the target at exceedingly high speeds and strike the metal plate with great force. This terrific bombardment causes some of the atoms of the target to have the electrons in their inner rings moved from one energy level to another and back again in such manner that X rays are generated and emitted by the atoms. It is possible to construct the filament housing in such manner that the electron stream is directed to a small point on the target and also possible to arrange the target so that the emitted X rays come out of the tube only in a given direction and thereby permit their application to any desired small area.

One of the most widely employed applications of the X rays is in the field of medicine. By means of these penetrating rays physicians and trained technicians are able to examine the structure and observe the functioning of the body's internal organs. Relatively small tubes which employ from 50,000 to 100,000 volts on the target anode are generally used for this purpose. These voltages produce rays sufficiently penetrating to pass through the body. The bones, teeth, and firmer fleshy parts, as well as foreign objects, obstruct their passage somewhat and cause them to cast a darker shadow than do the remainder of the tissues. Usual practice in making such examination is to place the patient between the X-ray tube and a large photographic plate; the rays thus make a shadowgraph picture of the exposed part visible when the plate is developed. Sometimes a fluorescent screen is substituted for the photographic plate, and the shadowgraph is visible immediately. X rays

constitute one of the most valuable tools ever developed for diagnosing disease of or injury to the internal parts of the body and for studying the progress of their recovery. The photograph at the beginning of this chapter is one made by a new ultrahigh-speed X-ray tube with which it is possible to get exposures as short as one-millionth of a second. In this action photograph of a football being kicked, all movement has been stopped, yet the details such as the bones in the kicker's foot and the nails in the heel and sole of the shoe are clear and distinct. High-speed X-ray photography will make possible the examination and study of many internal body movements which man has not been able to observe hitherto.

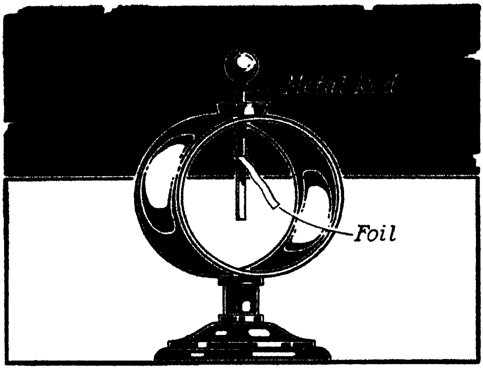
When great penetrating power is required, special X-ray tubes are used which operate at much higher voltages on the target than those mentioned above. The effect is to speed up the electrons striking the target and thereby produce X rays of shorter wave length. The shorter the wave length the greater the power of penetration. Waves of short wave length and high penetrating power are usually called "hard" X rays; those of longer wave length and less penetrating power, "soft" X rays. In the most powerful X-ray tubes in use at present, anode voltages run as high as from one to four million volts. Such tubes produce ultrashort X rays of a wave length about the same as the gamma rays of radium, so that these X rays not only have the power of penetrating tissue easily but also are able to penetrate several inches of steel.

The most important use of these extremely high-voltage tubes is to produce X rays of wave lengths that come within the wave-length range of gamma waves so that X-ray outfits may be used as a substitute for radium in the treatment of cancer and other diseases requiring radium treatment. Highly penetrating X rays are about as effective in killing diseased tissue as are the gamma rays from radium. As a matter of fact, X rays and gamma rays destroy good tissue in addition to their destructive action on diseased tissue, and it is essential that they be used with extreme care in the treatment of disease. The primary advantage of ultrashort X rays in the treatment of disease is that the tubes producing them may be made by man, and the desired radiation thereby made available to hospitals generally;



The world's largest X-ray machine to operate at 1,400,000 volts, showing complete assembly of main generator (center), voltage divider (left), and X-ray tube housing (right). (Courtesy of General Electric.)

whereas the only source of gamma waves is radium, and radium is a scarce and extremely costly natural material.



In a charged electroscope the foil strip is held away from the rod by the force of an electric charge.

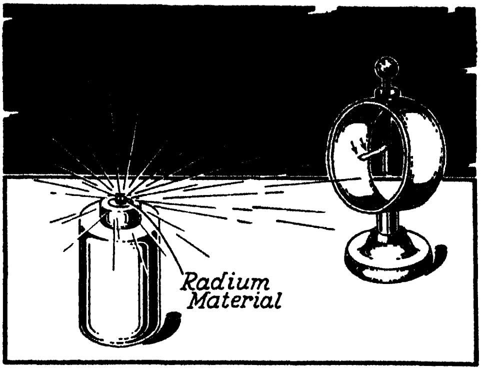
Probably the most important industrial application of X rays is in inspecting manufactured materials and other objects for hidden inner flaws. An X-ray photograph will usually reveal any imperfections. X rays may be used to examine the inside of locked or sealed packages, or packages suspected of concealing bombs by making shadowgraphs of the packages or examining them by means of the fluoroscope. They may be employed in inspecting radio tubes, golf balls, and other nontransparent objects.

Sir William Bragg and Henry Moseley of England were the first to use X rays for taking pictures of atomic structure and in the study of crystal formations. Since these rays are much shorter than visible light, it is possible to use them to observe by means of photographs much smaller objects than the highest power microscope can reveal to the eye. An X-ray photograph of a crystal structure would show a design that to the layman would not look much like a crystal, yet the specialist could use it to arrive at an understanding of the arrangement of the molecules in the crystal structure. Such pictures consist of interference patterns produced by molecules in the crystal acting like gratings through which the different waves pass. Since the molecular arrangement will determine exactly the type of gratings within the crystal structure and thereby the interference pattern produced, it may be readily understood that such a pattern reproduced on a photographic plate can be used to determine the crystal structure.

Gamma Radiation

Gamma rays are produced in nature when certain kinds of atoms explode and disintegrate; in this respect they might be

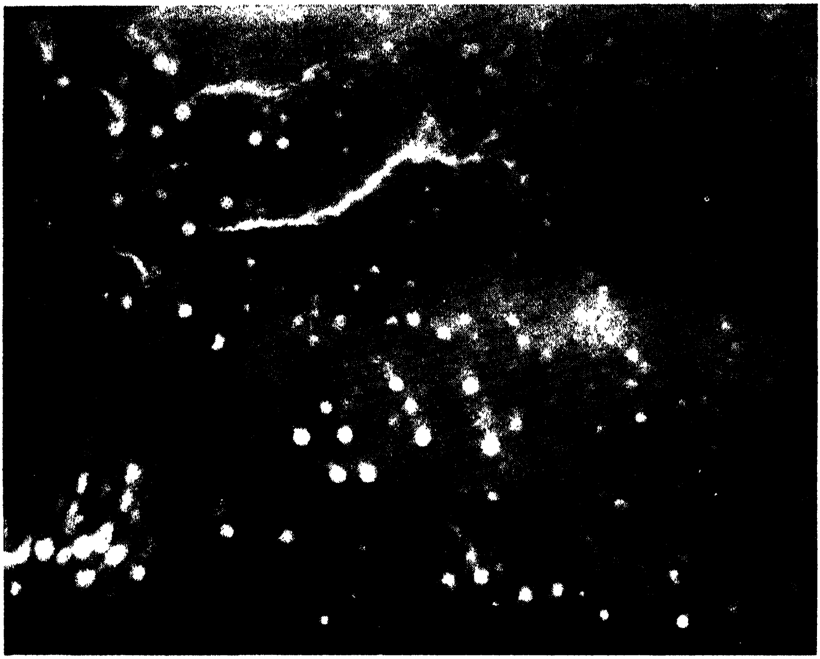
called the "death cry" of radioactive materials, such as radium, uranium, and others. It will be recalled from the discussion in an earlier chapter that when radioactive atoms disintegrate, two kinds of particles and one form of radiant energy are emitted. The particles are the positively charged alpha rays and the negatively charged beta rays; the radiant energy constitutes the gamma rays.



Rays from radioactive atoms ionize the air and permit the charges on an electroscope to escape, causing the foil to drop.

Shortly after their discovery by Becquerel in 1896, gamma rays were found to be a part of the electromagnetic spectrum adjacent to the X rays, and to range in wave length between approximately 1 angstrom unit and 0.01 angstrom unit. Even though they produce no visual or other sensations in the body, it is possible to detect their presence and to measure their strength. One way is to use a charged electroscope. This will retain its electric charge for a considerable time unless the charge is carried away by a conductor. Ionized air is such a conductor and it has the property of discharging an electroscope, as it permits the charges to leak off on to the ionized air particles. Gamma rays ionize the air through which they pass and therefore this property may be used to detect them and measure the strength of the rays. The electroscope becomes discharged in the presence of gamma rays, and the rate at which it discharges is a measure of the rays' intensity.

Gamma rays will produce fluorescence in certain chemical substances, and this property, too, may be used to detect them. Such chemicals shine with fluorescent light in the presence of gamma waves; in fact, this is the property which is made use of in luminous-face watches and clocks. The luminous paint consists essentially of a fluorescent material mixed with a small quantity of a radioactive substance that emits gamma rays. The fluorescence thus produced is strong enough to make the paint visible in the dark. A photographic plate is also affected



Gamma rays as well as hard X rays are used to find inner faults in metal castings. In this photograph, air bubbles and cracks are easily visible. (Science Service photograph.)

by gamma waves in the same manner as it is by light, ultraviolet, and X rays. The photographic technique may be used, therefore, to detect gamma rays and also to make shadow photographs of objects through which the rays pass.

An instrument known as the Geiger-Muller counter is the most sensitive apparatus for the detection and measurement of gamma rays. It is an electric device which amplifies the charge produced by the ionization of gas molecules in a cylinder when the gamma rays enter. The amplified charge may then be used to operate an electric counter or to produce a response in another measuring device. The instrument will respond to a single charged particle in the ionization chamber, and it is the most sensitive electric device that man has invented.

A number of valuable uses have been found for gamma waves, the most important of them probably in medicine where rays are employed in the treatment of cancer and other tumors. Radioactive material is usually placed inside a tiny sealed glass

capsule which may be inserted into the tumorous growth by a surgical operation. Gamma rays emitted by the radioactive material penetrate the tumorous tissue and destroy it. As these rays are destructive of good tissue also, and to a greater degree than X rays, extreme care must be exercised.

Because of their great penetrating power (much more than that of X rays), gamma rays have been employed to discover flaws in metals. In using them for this purpose a small capsule of highly concentrated radioactive material is placed on one side of the metal to be examined, and a photographic plate in a plate holder is placed on the other. The rays, after penetrating the metal, fall on the photographic plate and expose it. When the photograph is developed, any flaws inside the metal will show up as bright spots or as other irregularities in the picture. The penetration of the waves is in inverse proportion to the density of the metal and to the atomic weight of the element. Considerable thickness of steel may be penetrated by the gamma waves, but lead has such density and atomic weight as to prevent their passage to any marked extent. Lead is one of the best substances known for the absorption of these penetrating rays, and lead-impregnated cloth is used as a protective shield by people who work regularly with gamma and X rays.

Cosmic Rays

The first suspicion of the existence of other tremendously penetrating rays, now known as cosmic rays, was aroused soon after the discovery of X rays and gamma rays. During the first ten years of the twentieth century it became well known to investigators that even the most carefully insulated and screened electroscopes would discharge slowly, as if the surrounding air were always slightly ionized. This ionization of the atmosphere was first interpreted as coming from traces of radioactivity in rocks or other surrounding materials. Such interpretation was finally disproved by taking the screened electroscopes over lakes and up on high towers; always the slow discharge persisted. Finally, in 1913, the German scientist Hess carried an electroscope up to about 18,000 feet above sea level in a balloon. As the altitude increased he found a small but definite increase in the ionization. The results show that a radiation of very

great penetrating power was falling upon the atmosphere from above, and even at the lower levels it produced the ionization observed in the screened electroscopes. Thus it became established that the cosmic rays are coming into the earth from outer space.

Cosmic rays were first thought to be a form of electromagnetic radiation, and because of their great penetrating ability (approximately 100 feet of water or six feet of lead) they were considered to be shorter than gamma rays in wave length. Subsequent researches showed that some components were more penetrating than others and, furthermore, that some of the components were electrically charged particles rather than radiant energy. Yet it was so difficult to determine the nature of cosmic rays that a number of the most prominent investigators in this field maintained until recently that the most penetrating component was electromagnetic radiation similar to but shorter in wave length than gamma rays. It is now known, however, that cosmic rays are exclusively electrically charged particles. The only electromagnetic radiation associated with them is the possible secondary emission of gamma rays by atoms that may be disrupted upon impact with the incoming particles.

The charged particles that come into the earth from outer space, referred to as primary cosmic rays, pierce the outer atmosphere and collide with the atoms of the air. Out of these impacts, occurring at energies as high as twenty billion electron volts, which exceed by far any other energies known to man, come a series of atom-smashing changes that can be likened to the debris resulting from a bomb exploding within a building. This debris of smashed atoms constitutes the cosmic rays which reach or approach the earth's surface and which are subject to man's measurement. Such measurement has shown two general types of cosmic rays coming into the earth's lower atmosphere, the soft rays (most of which are absorbed before they penetrate to the surface) and the hard (or highly penetrating) rays. The soft components account for about 30 per cent of the total cosmic radiation measured and the hard rays for about 70 per cent.

The soft rays consist of a conglomeration of charged atomic particles. When the primary rays enter the upper atmosphere,

A stratosphere balloon being made ready at Ft. Meade, S.D., for ascent to make cosmic-ray measurements. (Photograph by courtesy of Major Lee Wells.)

electrons are stripped from atoms, and the electrons speed on to the earth's surface. Or, a primary cosmic ray may encounter an atom and send out from it radiant energy, and the energy may then be absorbed by another atom from which are ejected positrons and electrons. These, in turn, continue toward the earth's lower atmosphere. Then, again, primary cosmic rays may encounter and "explode" an atom of the upper atmosphere and send protons, positrons, and electrons speeding toward the earth. All of these secondary particles make up the soft components reaching the lower atmosphere.

The hard component of cosmic rays detected at sea level consists mostly of a charged particle only recently known to science, the mesotron. Created at high altitudes by primary cosmic rays encountering atoms of the air, mesotrons have a mass about 180 times that of an electron and may have either a

positive or a negative electric charge. They speed to the earth with energies as high as ten billion electron volts. In addition to their great energies and the dual nature of their electric charge, the mesotrons have another distinction in the realm of physics. They, alone among atomic particles, are unstable and "decay" after a half life of about one-millionth of a second. Some idea of their extreme velocity may be gained by knowing that they are born high in the stratosphere and yet can be detected at the earth's surface. Calculation puts the velocity required for the mesotrons to reach the earth before decaying at about 180,000 miles per second, which approaches the speed of light and constitutes the greatest velocity of a moving particle known to man.

Cosmic-ray investigation has been actively pursued during the last decade and huge sums of money have been invested in the researches. Out of these efforts has come a better understanding of the energies and materials of nature; and there is at present some promise that this knowledge may prove of immense practical value. Just recently it has been discovered by Prof. A. H. Compton and his associates at the University of Chicago that the intensity of cosmic rays as measured at the earth's surface is greatly affected by the same changes in the atmosphere that cause weather. Barometric pressure, temperature of the air, and its distribution in space overhead are known to produce variations in cosmic-ray intensity. Scientists now believe that if weather factors can produce such changes, it should be possible to reverse the situation and to use cosmic-ray variations to forecast the weather. To predict weather by cosmic rays is a long-range project which may require years to achieve. There are many complexities that must be overcome, but it appears that some day cosmic rays will be a useful tool for more accurate and longer range weather forecasting.

Infrared, the Warmth-giving Radiation

The electromagnetic waves that are just longer in wave length than those of visible light are the infrared rays. They are commonly called radiant heat or "heat waves," and they were first discovered as a distinct phenomenon by Sir William Herschel in the year 1800. They constitute the invisible energy

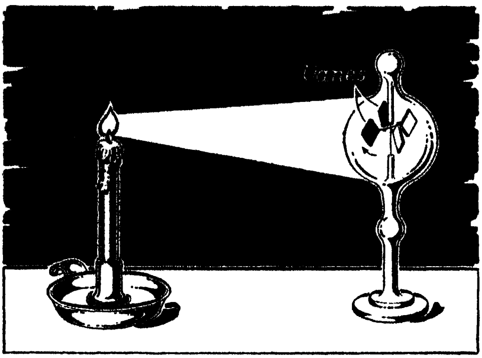
coming through space from the sun which is easily detected by one's skin on a hot, clear day. One can also feel the effects of this same kind of radiation on his hands when he holds them before a fire, since the burning fuel emits infrared rays. In fact any warm or hot body will emit these rays, and the amount of the radiation given off is proportional to the temperature and size of the body.

Infrared rays range in wave length from 7,800 angstrom units to about four millimeters. They are invisible, of course, but respond to optical instruments and standard lens practice much the same as do light waves. For example, infrared rays may be focused by the same kinds of lenses used for focusing light rays. The ancient "burning glass" was no more than a kind of lens that was used to bend the infrared rays of the sun to a focus at which the concentrated energy would raise the temperature of inflammable material to its ignition temperature. It is common knowledge that paper may be set on fire by using an ordinary reading-glass lens to focus the sun's rays on it. It would not be impossible for a person in an ice-covered country to start a fire on a clear day by shaping a lens out of ice and focusing the sun's rays on some easy-burning material.

Infrared rays of different wave lengths may be spread out into a spectrum, in a manner similar to light and with the same apparatus. When a lamp that emits both heat and light waves is mounted before a spectroscope, not only will a color spectrum of light be seen on the screen but also infrared rays may be detected just beyond the limits of the visible red. The red radiation merges into the infrared at about 7,800 angstrom units; the exact boundary line for any individual viewing the screen, however, will be determined by the limits of vision of his eyes. These invisible rays may be detected and measured in a number of ways, chief of which are by the sensation of heat in the skin, by a radiometer, and by the thermocouple. Special photographic plates will also respond to certain infrared radiations.

The radiometer consists of four balanced vanes mounted on a vertical axis in a glass tube filled with air at low pressure. Alternate sides of these vanes are covered with a black substance which absorbs heat rays, and the remaining sides are polished metal which reflects the waves. When infrared radiation is

absorbed by the black sides of the vanes, the temperature is raised slightly, and a higher gas pressure is created there. Owing to the increased pressure, the vanes rotate in a direction away from the black sides, and the greater the amount of infrared energy falling on the vanes the faster will be the rotation. The rotation of the radiometer vanes when the instrument is placed in a show window beneath a lamp or in the sunlight is a scene familiar to many people. The instrument shows speedy rotation, also, when placed just beyond the red part of the visible spectrum of the sun or a lamp source, showing that infrared rays are present at that position in the spectrum. This was the astronomer Herschel's discovery and the one that first established the idea that heat and light waves were the same kinds of radiation.

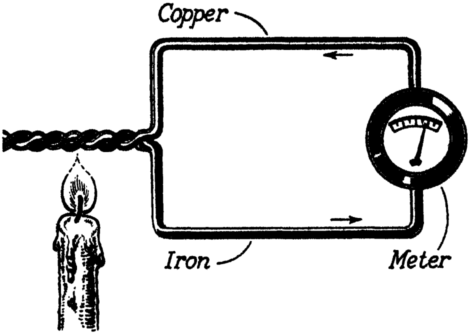


Radiometer.

The thermocouple is an even more accurate detector and measuring device of heat waves and the standard scientific instrument used for this purpose. It consists of a junction of two different metals, such as iron and copper wires. The junction is arranged in such manner that the radiation falls on the connection only; and when this happens, a small electric potential is set up in the wires. Should the other two ends of the wires be connected through a galvanometer, an electric current will flow through the system, and the galvanometer will register it. The galvanometer may then be calibrated so as to measure the radiant heat, since the electric current generated will be in proportion to the amount of radiation falling upon the junction.

With such an accurate measuring device, it is possible to determine exactly the amount of heat radiated by hot bodies. This has been of wide practical significance in many ways. It has permitted a correct determination of the amount of heat radiated by the sun, and its effect on weather conditions. It has been of value in designing home and office radiators so as

to get an increased value for the fuel consumed. The electric-lamp industry is vitally concerned with the development of lamps that will convert less of the electric energy into radiant heat and more of it into light. Measurement of these energies is the first prerequisite, of course, for such development.



Thermocouple.

Within recent years special photographic plates have been made that are sensitive to infrared rays. Although infrared photographic technique is not used primarily to detect and measure this radiation, it has found important applications in other instances. By using these special plates, pictures can be made of warm objects in fog or in complete darkness. Furthermore, infrared rays penetrate the air and fog much better than light waves do, and the same is true of the way in which they penetrate the skin of the body and many pigments and paints. Infrared photographs of distant objects and scenes show much greater detail than do pictures of these scenes made with light. Photographs of the surface of the body made with infrared radiation will often show causes of disturbances that are difficult to detect otherwise.

Infrared rays are also becoming important in navigation on sea and land in foggy weather and at night. Emitted by ships or other conveyances, they cut through fog or darkness without marked absorption and may be detected with a thermocouple. By noting the strength of the current as the thermocouple is rotated in different directions, it is possible to determine the direction and distance of invisible ships at sea or of boats in a crowded harbor. Radiant heat has found application in medicine as a substitute for the old-fashioned poultice or hot-water bottle, as attested by the fact that infrared lamps are widely sold. The lamp is placed near the body of the patient, and its radiation allowed to fall on the skin at the place where pain is felt. Heat rays often remove pain, and it may be possible that

these heat rays penetrate more deeply than the mere contact heat of the earlier and simpler appliances. A hot-water bottle does emit, of course, infrared radiation but because its temperature is much lower than that of an infrared lamp, the rays are much longer in wave length and have little penetration.

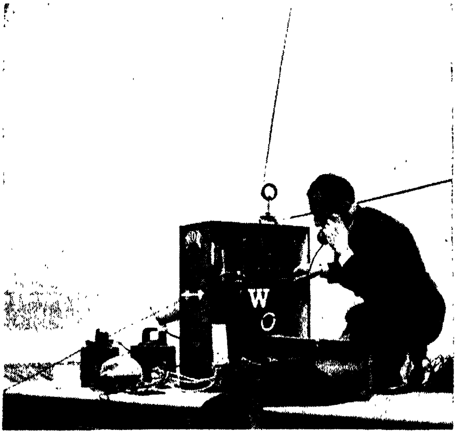
Radio Waves, the Messenger Radiation

The longest waves of all the invisible energies of the electromagnetic series of radiation are the radio waves. The possibility of the existence of waves of wave length longer than infrared was first suggested by the work of Sir Isaac Newton and Michael Faraday of England and was first experimentally established by Heinrich Hertz of Germany in 1887. They were called for a time the "Hertzian waves." The Italian scientist Guglielmo Marconi was the first to use them, in 1896, for communication over a very short distance. Since that time the use of radio waves, as they are now called, for communication has extended to the limits of the earth.

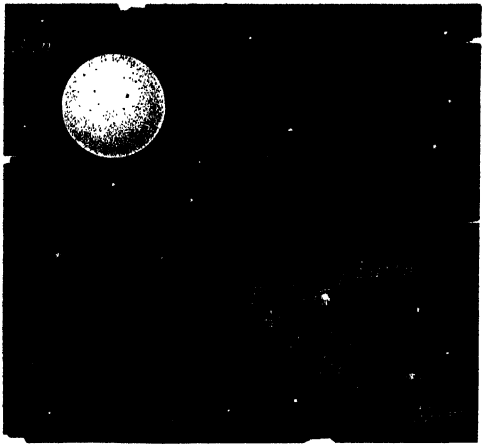
Radio waves have a wave length ranging between 2.5 millimeters and about thirty thousand meters. The longest are, therefore, several miles in length, and the shortest overlap the longest infrared rays of about 4 millimeters produced by heating an object. No sharp line of separation exists, then, between the longest infrared rays and the shortest radio waves. The latter vary considerably in their power of penetration of the air and in their mode of travel through the earth's atmosphere. For this reason they have been divided into several groups, or divisions, of the radio spectrum, depending on their wave length.

The shortest group are the so-called microwaves, which range from a fraction of a centimeter to one meter in length. These will not travel around the earth's curvature, since, by traveling in a straight line, they penetrate the entire atmosphere and are radiated out into space. They have, therefore, short range along the earth's surface and are used only for local transmission of messages. They are very similar to light waves in their transmission characteristics, and for their most effective handling it is necessary to employ the proper sort of "lenses" and reflectors to direct them into parabolic beams.

A microwave projector containing a parabolic reflector, aboard the S. S. Normandie. (Photograph by Herman Young.)



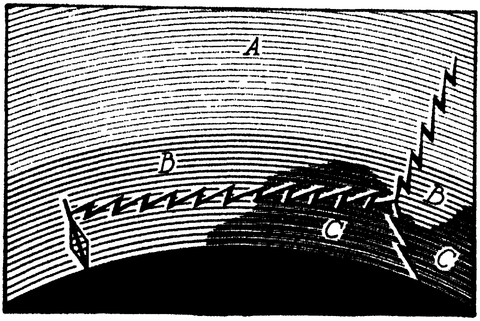
An ultrashort-wave portable transmitter and receiver in operation. The small vertical rod is the antenna. (L. M. Cockaday photograph.)



Moon and sun's effect in producing high tides in the earth's air which influence the reflection of radio waves.

The next group are the ultrashort waves which range in wave length from one to ten meters. They retain most of the optical characteristics of the microwaves of traveling through the entire atmosphere and straight out into space, but under favorable atmospheric conditions they are able to travel around the curvature of the earth to considerable distances. This is particularly true of the longer ones of five to ten meters wave length. In 1938, Lieut. Comdr. L. M. Cockaday of the United States Navy discovered that waves as short as four to five meters could travel two or three thousand miles over the earth's surface, under favorable atmospheric conditions, especially during spring, summer, and fall weather.

Research conducted since that time shows that ultrashort waves can be and have been transmitted around the earth to distances of 3,000 to 4,000 miles by being reflected from a highly conductive layer of atmosphere that forms as high as 200 miles above sea level on certain occasions and acts as a mirror to send the waves back to the earth. These periods occur regularly at the time of full moon. The explanation may be visualized by referring to the accompanying diagram. The atmosphere is evidently pulled up to great tides by the moon's and sun's gravitational attraction, not unlike the way in which the tides rise on the surface waters of the earth. At the time of full moon, the sun, earth, and moon are in line, and the gravitational



Radio waves bend by refraction.

attractions of the sun and the moon on the earth's waters are combined. At this time the ocean tides rise higher than at other times, as represented in greatly exaggerated scale at A and A_1 in the drawing. The same condition holds true for the earth's atmosphere, except to a greater extent, as shown at B and B_1. During this time there is a much greater area of atmosphere between A and B and between A_1 and B_1 than at other times during the month, so that the ionized conducting area itself may be very thick and high. The ultrashort waves under these occasional conditions may be reflected from this highly ionized or conducting layer and come down to earth at a distant point. Under ordinary conditions of lesser thickness of the ionized portion of the atmosphere the ultrashort waves pass through the atmosphere and never come down to earth again. In this respect their behavior closely parallels that of the microwaves, which always escape beyond our atmosphere.

Another condition under which ultrashort waves may be transmitted to distances of 100 to 300 miles is caused by refraction, as illustrated in the above diagram. When a weather cyclone brings cold air into an area that has been enjoying warm-weather conditions, the colder, denser air pushes underneath the warmer air. At the junction of the cold and warm layers refraction and some reflection of the ultrashort radio waves take place, bending some of the radio waves down to earth again where they can be received. Temperature inversions of this type are found quite frequently, and transmission from 100 to 300 miles can be accomplished once or twice each week with ultrashort waves because of these conditions.

Still another unusual property of the ultrashort waves is that reflection from the ground may produce strong signals at a considerable distance and at high altitude. If an airplane is 100 to 200 miles distant from the transmitter at a position similar to the point marked X in the accompanying drawing, reflection from the earth at point A may bring the signals up to a high strength so that they may be clearly received there even though the transmitter power is very low. It is not impossible that this property may cause the ultrashort waves to become an important factor in aviation radio of the future.

Reflection of radio waves from the ground.

Adjacent to this band is the so-called short-wave group, which ranges in wave length from 10 to 165 meters. It is noted for traveling around the earth to the greatest distances with the least amount of power and is used extensively in short-wave broadcasting. The "short-wave" reception familiar to most laymen is broadcast on frequencies within this band.

The medium, or critical, wave group ranges from 165 to 2,600 meters in wave length, and these wave lengths are the ones used for local broadcasting in America as well as abroad. In America waves between 165 and 550 meters only are used; in Europe the range extends to 2,600 meters. These wave lengths cover service areas of a few hundred miles more consistently than do other wave lengths. They are heard at much greater distances, however, occasionally.

For high-power communication the long-wave group ranging in wave length from 2,600 to 30,000 meters is employed. Wave lengths longer than this could be produced by ordinary alternating currents, such as that used for lighting and heating homes; however, the antennas to radiate them would be so huge that they would be prohibitive in cost and almost impossible of construction. Wave lengths most used for long-distance, long-wave-length, commercial communication are those around 10,000 to 15,000 meters.

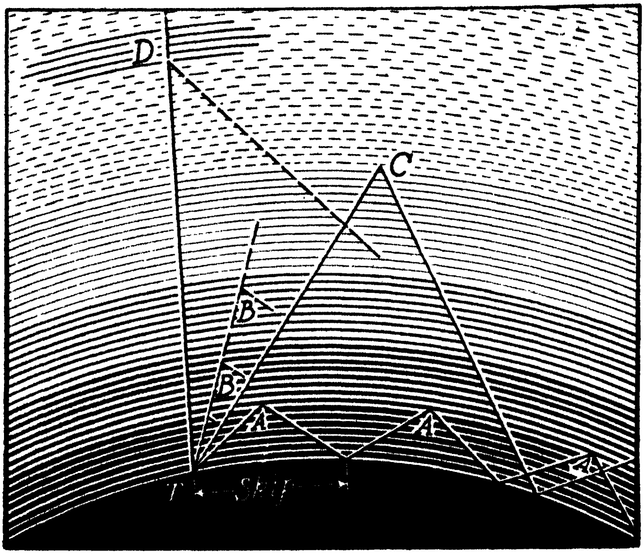
How Radio Waves Circle the Globe

Because radio waves, like all other electromagnetic waves, travel in straight lines away from the antenna source, it was at first predicted that because of the earth's curvature they could not be used for communication over long distances. Later experiments showed, however, that they could travel around this curvature, and the discovery of how this is accomplished constitutes one of the most interesting and important achievements in radio research.

An eccentric English recluse and mathematical physicist, Oliver Heaviside, had at the beginning of the twentieth century conceived of a condition in the upper atmosphere that would permit radio waves to travel around the earth. He had calculated that there was an upper ionized layer of the earth's atmosphere. As reports of his mathematical theory were circulated, others reasoned that such a layer might reflect radio waves back to the earth at distant points. Even though Heaviside objected so strenuously to outside human contacts that he had his meals brought to him by a London policeman, he was finally persuaded to come to America to assist Prof. Arthur Kennelly of Harvard University with the experimental verification of his mathematical data. Together they established the presence of such an ionized layer in the upper air, and it has been called, therefore, the Heaviside-Kennelly layer.

Since that time a number of layers of ionization high up in the atmosphere have been verified, and the region is generally referred to as the ionosphere. Much experimental evidence of recent date indicates that there are as many as six reflecting layers one above another in the ionosphere, and even higher layers are suspected to exist at certain times.

These ionized layers are formed in the extremely low-pressure areas of the upper atmosphere. In such upper reaches there are no winds or temperature gradient and little movement of the air; also, it is here that the various gases of the atmosphere divide themselves into layers, with the lightest ones at the top. The various gases in the outer layers from 125 to 190 miles high are ionized by the strong ultraviolet rays and other radiations of the sun. The lower layers, 50 to 60 miles in



Transmission of radio waves of different wave lengths from transmitter T around the earth is brought about by reflection from different ionized layers of the atmosphere.

height, are believed to be ionized by corpuscular rays from the sun and perhaps by cosmic rays from outer space. All these layers or masses of ionized particles reflect radio waves somewhat as glass mirrors reflect light. Radio waves going upward and impinging on them are reflected to the ground at a distant point around the earth's curvature, as shown in the drawing.

The long-wave radio waves, designated by *A* in the drawing, because of their lack of penetration are reflected to the earth from the lower layers to a point some distance away. Between the transmitting antenna and this point where the waves bound back to the earth is an area called the "skip distance" where little or no energy is received. Within the skip distance a radio receiver would not be able to receive much energy from the transmitter and, therefore, could not satisfactorily receive broadcasts, although a receiver farther away would receive the transmission with a strong signal. Beyond the point at which the waves come back to the earth, they are reflected from the earth's surface to the Heaviside layer again, then back to the earth, and so on around most of the earth's curved surface.

Radio waves of medium wave length, such as those used for regular broadcasting in the United States, have more penetration and pass through the lower layers with some absorption. These are the critical wave lengths which are mostly absorbed as they pass through each layer, with only slight reflection to earth, as is shown at *B* in the drawing. They are not very useful, therefore, in long-distance communication but are employed for local broadcasting.

Short waves, however, have sufficient penetration to pass through the ionized regions of the lower atmosphere and are then reflected to earth from strongly ionized regions at much higher altitudes in large jumps, so that they cover great distances with little loss of energy. These waves are represented at *C* in the drawing. The ionized layers vary in their density and also in their distance from the earth. Variations occur daily with daylight and darkness and also yearly from summer to winter. This effect makes some wave lengths more efficient than others at certain times of the day and night for any desired distance range, and the same holds true for summer and winter. Daylight and summer transmission seems to be more favorable for the shorter short waves, whereas nighttime and winter conditions favor the longer short waves. This is the reason why foreign short-wave broadcasting stations change their wave length during the day, gradually shifting to the longer of the short waves as night comes on. In summer the shorter wave lengths are more favorable for the greater part of the day, but in winter the longer of the short waves remain more favorable over a greater portion of the day.

The ultrashort waves have such great power of penetration of the atmosphere that they usually pass through all or almost all the layers without reflection. These are the waves that would have to be used to communicate with other planets, should man ever succeed in so communicating. During the last few years many experiments have been conducted to show that even five-meter ultrashort waves may be reflected to the earth under favorable atmospheric conditions. One theory to explain the fact that these waves are received at times over quite long distances presupposes an occasional extremely dense reflecting layer. Such a layer may consist of a dense cloud of electrons, or

possibly ionized particles, from the sun, deflected by the tidal forces of the moon and concentrated in a small area high in the atmosphere. This layer when present reflects the ultrashort waves as shown at *D* of the drawing. The ultrashort wave band, however, is not regularly reliable for distant communication. The greatest long-distance records for this band have been obtained during full-moon periods in the summer months when the sun's ionizing power and its corpuscular radiation are the strongest.

Radio waves, besides being employed for radio telegraphy and radio telephony, are also used for transmission of photographs by the facsimile process from one place to another over long distances. Because of lack of reflection and fading, the shorter radio waves are now found to be the most applicable to television transmission, and this is especially true of the shorter waves below four meters.

Ultrashort waves are also used to a certain extent in medicine, to produce fever in local parts of the body to which they are applied. By proper adjustment of the antennas, the ultrashort radio waves may be made to penetrate deeply and to set up strong heating currents in a local area, raising the temperature of that portion of the body well above 100°F. In this way an artificial fever can be produced which is instantly controllable and also localized so that the offending cause may be destroyed without affecting the general temperature of the whole body system.

Electrical Source of All Radiation

The six groups of electromagnetic waves, as discussed in this and the preceding chapter, extend in wave length from billionths of an inch to miles. Except for this wide range in wave lengths and a wide variation in their power of penetrating matter, the six divisions share many properties in common. Such similar characteristics indicate a common type of origin for the entire electromagnetic spectrum. A clue to the electrical character of the sources of these waves is gained from the method used to generate radio waves. They are produced by moving electric charges forward and backward in an antenna. The rhythmic movement of the charges generates a corresponding

sequence of radio waves. Likewise, the other divisions of this radiation family are generated by oscillating electric charges.

It is now known that when large units of electric charges are moved relatively long distances, long waves of low penetrating properties are produced; and that when small units of electric charges are moved short distances, short waves of great penetrating qualities are generated. Radio waves are produced by moving large masses of electrons back and forth in great broadcasting antennas. Infrared rays are caused by electrons oscillating within the atoms of heated substances. Visible light, ultraviolet waves, and X rays are generated by electrons moving shorter distances, from one energy level to another within the atoms, thus producing shorter waves. Gamma rays probably are caused by electric changes taking place within the very nuclei of atoms or by the formation or destruction of atomic nuclei and, therefore, involve small movements of electric charges which produce these pigmy waves of nature.

REFERENCES FOR MORE EXTENDED READING

GAMOW, GEORGE: "Mr. Thompkins in Wonderland," The Macmillan Company, New York, 1939.

A renowned physicist writes a fantasy on a man transported to a world in which the laws of atomic physics affect large-scale objects such as man and machines.

GRAY, G. W.: "The Advancing Front of Science," Whittlesey House, McGraw-Hill Book Company, Inc., New York, 1937.

Chapter V of this delightfully written account of some recent scientific achievements explains the use of infrared photography in aiding man to "see" things that are otherwise invisible.

WATSON, H. M., H. E. WELCH, and G. S. EBY: "Understanding Radio," McGraw-Hill Book Company, Inc., New York, 1940.

The authors have produced here a very practical account of the fundamentals of radio which should be useful for home study.

MILLIKAN, R. A.: "Cosmic Rays," The Macmillan Company, New York, 1939.

The man who has been extensively and intimately connected with cosmic-ray study since the discovery of these rays has combined into a book three lectures delivered at Trinity College, Dublin. This book is an account in the author's interesting and fast moving style of the characteristics and properties of cosmic rays and his interpretation of their nature and source.

CLARK, G. L.: "Applied X-rays," 3d ed., McGraw-Hill Book Company, Inc., New York, 1940.

A thorough and comprehensive treatment of X rays and their applications. The material is suitable for advanced students in this special field, but parts of it are useful to the layman desiring specific information.

ELLIS, C., A. A. WELLS, and F. F. HEYROTH: "The Chemical Action of Ultraviolet Rays," 2d rev. ed., Reinhold Publishing Corporation, New York, 1940.

An extended and thorough treatment of the subject but valuable as a reference for the layman wishing to secure specific information.

Science and Society, published by Science and Society, Inc., New York.

A quarterly journal containing many well-written and thought-provoking articles on the relationship of science to modern life, industry, and the social order.

General Electric Review, published by General Electric Company, Schenectady, New York.

Popularized and technical articles on all phases of the electrical industry, illustrated with many remarkable photographs.

General Electric.

13: MOVING ELECTRONS

Or the Science of Electrical Engineering

ELECTRICITY is not a recent discovery. If one delves into its history, he will find that the Greeks knew about it over two thousand years ago. To be sure, their knowledge did not extend far, but it was fundamental. They discovered that rubbing a piece of amber with cloth made it attract other objects; small bits of fiber or straw would actually fly to it and cling to its surface. The same phenomenon can be observed by anyone who combs his hair vigorously with a hard-rubber comb on a cold dry day. The comb snaps and crackles and attracts individual strands of hair. We explain this effect by saying that the comb has become "electrified," a term that comes from the Latin word for amber. Many substances exhibit the property of becoming electrified or "charged" when rubbed;

and the fact that electrical charges can be produced on matter in this way is one of the most fundamental principles of electricity.


Ancient and modern examples of electric charges.

At another early period in human history, the Chinese observed that some kinds of stone could be used to determine geographical directions. An oblong piece of this stone suspended from the middle, as shown in the drawing, would gradually rotate in a horizontal plane until one end pointed toward the north and the other toward the south. This direction-indicating property was useful for both navigation and land travel in ancient times. Therefore these stones were called "leading stones," a term from which our word loadstone was derived.

We now know that these stones were iron ore, consisting principally of the black iron oxide, magnetite. The property of pointing in a north and south direction, long since referred to as magnetism, is explained by the fact that the earth itself is a large magnet with its magnetic force concentrated near the geographic poles. This force magnetizes the iron ore so that one end is attracted toward the earth's south pole and the other toward the north. Modern magnetic compasses make use of this principle by suspending a small steel magnet, called the needle, so that it can rotate in a horizontal plane. The earth's magnetic force keeps this needle pointing in approximately a north and south direction at all times.

These two ancient and apparently unrelated facts about electricity and magnetism form the basis for all of our electrical wonders of the twentieth century. Modern achievements with electricity, however, were not made possible by a mere knowledge of these two facts; it was necessary for someone to discover how to combine them. This combination opened up a multitude of possibilities which have not been completely explored even to the present time.

Electrified Matter

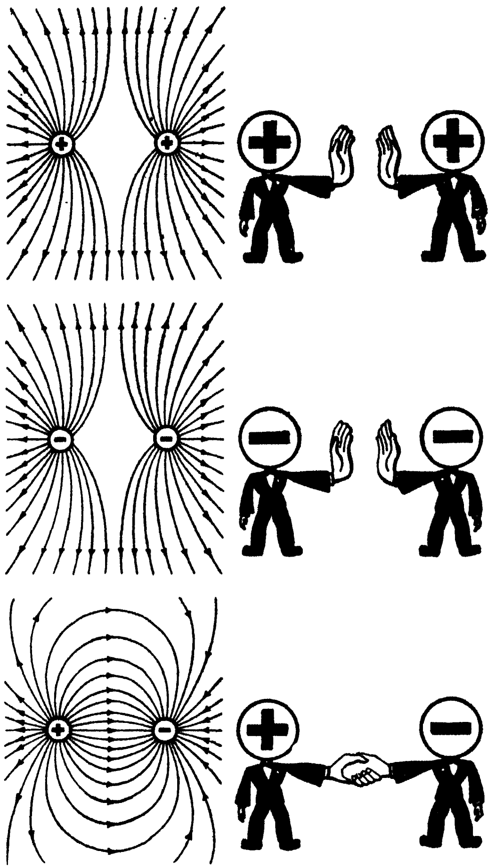
Over three hundred years ago it was known that a great many other substances possessed the same property that the ancient Greeks had observed in amber. Materials such as glass, sulphur, and sealing wax could be electrified by friction. Objects in this electrified condition would exert an attractive force upon other objects, and accordingly they were said to possess an electric charge.

Ancient and modern observations of the effects of magnetism.

It remained for research work of the past few decades, however, to explain the exact nature of an electric charge and show why some materials exhibit it when rubbed, whereas others do not. The answer comes from our present knowledge of the structure of matter.

We know that the atoms of all substances contain protons and electrons. The exact structure of these tiny particles is not yet fully understood, but that is unimportant for the purpose at hand. We do know that protons and electrons represent small charges of positive and negative electricity, respectively. We are aware of the charged condition of these particles because they exert forces of attraction and repulsion upon neighboring electrons and protons. Perhaps the most mysterious fact about these forces is that they require no tangible medium through which to act, since the forces between electrons and protons are exerted through a vacuum as well as through matter.

For convenience, the charge of the proton is designated by a plus sign; that of the electron, by a minus sign. A charged particle always exerts a repellent force upon charges of opposite sign. These laws of attraction and repulsion between electric charges are illustrated in the accompanying drawing. The lines shown between the charges in the diagrams have been placed there merely to show the direction in which the force is acting



Like charges repel and unlike attract.

in each case. This is the conventional way of representing the force between charges, and the lines are called lines of force. It is important to remember, however, that there is no physical equivalent of these lines in the case of the actual charges. The pictorial sketches to the right of each diagram are merely for the purpose of illustration.

The magnitude, or strength, of the force between charges decreases as the distance between them increases. As a matter of fact, the force decreases in approximate proportion to the square of the distance. Thus two charges separated by a relatively long distance will exert very little force upon each other, but two charges relatively close together will act upon each

other with a strong force. The region around a charge that is under the influence of its force is called the "electric field" of the charge. The lines of force shown in the drawing represent the electric fields around the charges. Our proof that such a field of force exists is the fact that one charge can affect another at a distance.

When an electron and a proton come very close together, their respective fields neutralize each other and produce a condition in which the combination outwardly exhibits no electric properties. Such a condition also exists within the atoms of normal unelectrified matter. Both electrons and protons are present but in equal amounts, so that outwardly the body is electrically neutral and is not surrounded by an electric field.

An electric charge can be produced upon any material body by temporarily removing electrons from a few of its atoms. Some substances can be thus charged more readily than others. When a glass rod is rubbed vigorously with a piece of silk cloth, the close contact between the two materials apparently transfers electrons from the glass to the silk. The glass, having lost some of its negative charge, is then surrounded by an unbalanced electric field and exhibits a positive charge. Similarly, fur or wool will deposit electrons on some materials, such as hard rubber or amber. A rubber rod that has been rubbed with a piece of cat's fur exhibits a strong negative charge, because it has gained electrons from the fur; similarly, a hard-rubber comb collects electrons from frictional contact with the hair. Thus we see that matter can be electrified by any means that will produce an uneven distribution of its electrons. At every point where there is either an excess or a deficiency of electrons we have the condition known as an electric charge.

In an uncharged body the electrons are held in place within the atoms by the attractive force of the protons. To charge a body, therefore, one must pull the electrons away from their normal positions within the atom. When displaced in this manner, the electrons behave as if they were connected to the protons by stretched rubber bands which tend to pull them back into place the instant that they are released. This "stretching" force represents electric energy which, if harnessed, can be made to do useful work as the electrons find their way back

to the protons. The entire science of electrical engineering is concerned merely with methods of utilizing this energy by controlling the paths, or circuits, over which electrons flow.

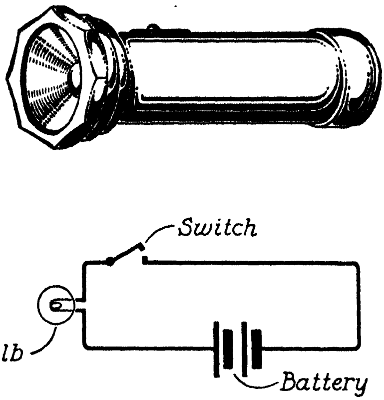
Electric Energy from Chemical Action

The production of electric charges by rubbing substances together was an interesting pastime for the early experimenters; however, this method was far too inefficient to suggest the possibility of using electricity for anything except a scientific curiosity. For practical use, any source of electric energy must obviously be arranged so that it will remove electrons from atoms continuously, since the utilization of electric energy requires that the electrons be returned to the atoms. A means must be provided, therefore for continuously replenishing the supply of moving electrons.

During the eighteenth century two Italian experimenters, Luigi Galvani and Count Alessandro Volta, discovered that continuous electric charges were produced on strips of dissimilar metals placed in a chemical solution. One metal strip became positively charged, and the other negatively. Certain pairs of dissimilar metals produced stronger charges than others. They also observed that if the metals were placed in contact at one end, some kind of disturbance took place which acted like a "flow" of electric charge through the metals. This disturbance was called a "galvanic current" and was designated by that name for more than a century thereafter. Now we call it an electric current.

The apparatus used by Galvani and Volta was the forerunner of the modern electric battery, and we now know that the currents that they observed were nothing more than a flow of electrons from the negative strip of metal in the chemical solution to the positive one and back into the solution. The modern battery, such as the one used in a flashlight, consists of a piece of zinc and a piece of carbon both in contact with a chemical solution in the form of a moist paste. The chemical reaction between the carbon and zinc, through the medium of the paste, removes electrons from the carbon and piles them up on the zinc. The positive and negative terminals of the battery are connected to the carbon and zinc, respectively. When the flashlight

is switched on, the terminals connect with the two ends of the fine wire filament in the flashlight bulb; the electrons rush from the zinc to the carbon through this slender filament and heat the filament so that it glows with light.



The electrical elements of a flashlight.

All modern electric batteries produce electric energy through chemical reaction. In some types, such as the flashlight battery, the reaction cannot be reversed. When the materials in the battery have completed their chemical change, the battery is "dead" and must be replaced. In others, such as automobile batteries, reaction is reversible; accordingly, when the battery has been "discharged," it is possible to "recharge" it by forcing electrons through it in the opposite direction.

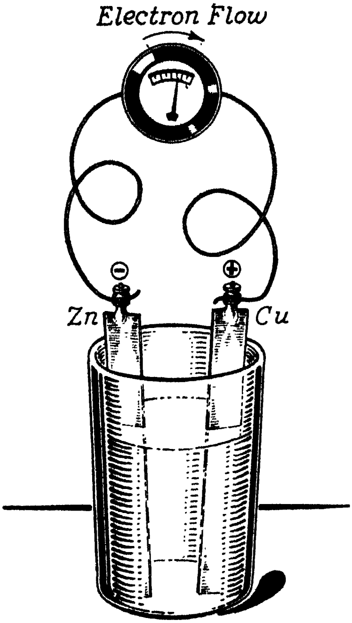
Galvani and Volta were unable, of course, to explain their findings in modern terms of chemical reaction. Their discoveries, however, made it possible for later experimenters to build electric batteries and experiment with electric currents, and from such experiments the real explanation for electrical mysteries began to come to light.

Currents Move Compasses

Around the beginning of the nineteenth century, electric batteries were the latest addition to the scientific equipment of most laboratories. They were generally known as galvanic or voltaic batteries and consisted of positive and negative electrodes immersed in a chemical solution. Copper and zinc were generally used for the electrodes, and the solution was either a dilute acid or salt water. A battery of this general type is illustrated in the drawing on the following page.

With the aid of these new devices, certain elementary facts about electricity were discovered. It became known, for example, that electric currents would flow readily through some materials,

whereas other substances seemed impervious to them. The materials that offered the least resistance to the flow of electric currents were said to be good conductors, the most typical of which are metals, carbon, and salt solutions. Substances that would not transmit electric currents, such as glass and rubber, were called nonconductors, or insulators. Also, it had been observed that the passage of electricity through a conductor has a tendency to generate heat in the conductor. Electricity, at this period, was considered an interesting scientific fact, but no practical uses for it had been discovered.



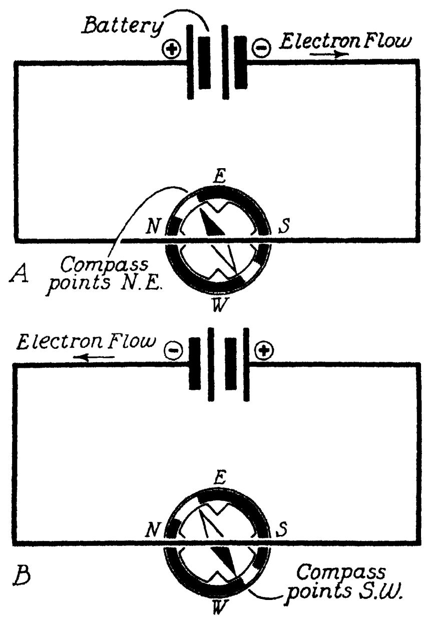
A modern example of the battery of Galvani and Volta.

In 1819 a Danish physicist named Hans Christian Oersted was experimenting with the effects of electric currents in a wire connected between the terminals of a galvanic battery. In the course of his experiments it happened that an ordinary magnetic compass had been placed on the same table with the battery and wire, somewhat as shown in the accompanying illustration. To his astonishment Oersted observed that when current was flowing in the wire, the compass needle no longer pointed in a north and south direction! Instead, the needle pointed nearly at right angles to the wire. When the connection to the battery was reversed, so that the current flowed in the opposite direction in the wire, the compass needle would reverse its direction; but the axis of the needle was still almost perpendicular to the wire, as shown in the lower drawing. Here was a mysterious relationship between electricity and magnetism that later proved to be one of the most important scientific facts ever discovered.

Oersted immediately made a series of additional tests which conclusively proved two important principles. The first can be summarized in the statement that an electric charge in motion

produces a magnetic force, technically called a magnetic field; this is the force that rotated the compass. The second principle is that the "direction" of the magnetic field depends upon the direction of flow of the electric charge. By direction of a magnetic field we mean the direction in which the north pole of a magnetic compass points when placed in the field. This direction of the force is what caused the compass to point in opposite directions when the current was reversed.

In Oersted's time, nothing was known about the existence of electrons, so he could not explain his results in terms of these units. However, the important point was the fact that electricity in motion produces magnetism. This relationship was named the "Oersted effect" and has been known by that name ever since. In later years knowledge of this effect made it possible to harness moving electrons, so that now they are used to do much of the work of mankind.



Illustrating Oersted's experiment.

Electricity and Motion Are Interchangeable

One of Oersted's most brilliant contemporaries was an Englishman named Michael Faraday. Faraday began his career as a bookbinder's apprentice and acquired a knowledge of science by reading the books brought to him for binding. During his early youth, he had an opportunity to listen to four lectures on chemistry given by Sir Humphrey Davy, at that time director of the Royal Institution in London. Faraday kept a neat and complete set of notes on these lectures, and when Davy later saw them he gave Faraday a job as laboratory assistant. Nothing could have pleased the young man more, and perhaps

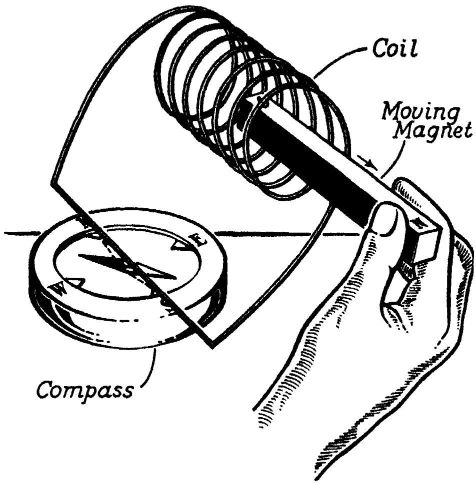
nothing could have been more fortunate for the immediate future of electricity.

While at the Royal Institution, Faraday heard of Oersted's new discoveries about the effects of an electric current on a compass needle. He became so deeply interested that he built apparatus similar to that used by Oersted and, for his own satisfaction, proceeded to duplicate all Oersted's experiments. Then he began to devise experiments of his own in an effort to find a means of producing continuous motion by means of an electric current. After long and careful experimentation on this problem, Faraday finally succeeded in solving it. His apparatus was simple and somewhat crude, but it did produce continuous mechanical motion from the interaction between an electric current and a magnetic field. Thus the very first electric motor was born.

His success with the electric motor inspired Faraday to further investigations into the mysteries of electricity and magnetism. Since these two things were so closely related, he believed it should be possible to produce an electric current in a wire by means of a magnet instead of with batteries. A memorandum in his notebook, made shortly after his invention of the motor, reads: "Convert magnetism into electricity." Thus Faraday assigned himself one of the biggest tasks in the history of electrical engineering, and he did it as modestly as the average citizen makes a note of a luncheon engagement.

The assignment proved to be difficult, and anyone possessing less perseverance probably would have abandoned it as an impossibility; in fact, many others of Faraday's time had tackled the problem and failed. For eight years he kept on patiently seeking a means of making magnetism produce a flow of current in a wire. Without benefit of our modern knowledge of electrons and the true nature of magnets, all his experiments consisted of attempts to make magnetism "flow" into a wire circuit by direct contact between a magnet and a coil of wire. He discovered, however, that it is not possible to convert magnetism into electricity in this manner.

One of the pieces of apparatus that Faraday built for his experiments consisted of a long coil of wire with the ends joined together in a loop, as shown in the drawing. A compass



How Faraday discovered electromagnetic induction.

was placed under the loop to indicate any possible flow of current. The experimental procedure consisted of placing various types of magnets inside the coil in an attempt to generate electricity from magnetism by contact between the magnet and the wire. One day as Faraday hurriedly slipped a magnet into the coil, he noticed that the compass needle moved momentarily but settled back to its original position as soon as the magnet ceased moving. This slight bit of activity did not escape his observing eye. He started to withdraw the magnet in order to repeat the experiment; and as he pulled it out of the coil, the needle moved again but this time in the opposite direction. Here, at last, was a clue to the way to produce electricity from magnetism! Apparently it was the motion of the magnet, rather than its mere presence, that caused a current to flow in the wire.

To test this reasoning he jiggled the magnet back and forth inside the coil. The compass needle responded by jiggling back and forth in a similar manner. Next he tried holding the magnet still while he jiggled the coil, and the result was the same. Now he began to understand why all his earlier experiments had failed, for previously he had kept both magnet and wire stationary. This latest experiment clearly showed that there must be relative motion between the magnet and the wire in order

to induce a current to flow in the latter. Instead of converting magnetism into electricity, Faraday had actually succeeded in converting motion into electricity through the medium of magnetism. Magnetism, to be sure, was an essential agent, but it remained unchanged during the process.

The principle of producing electric current by relative motion between a wire and a magnet is now known as the principle of electromagnetic induction, so called because a current is "induced" in the wire. It is perhaps the most important principle in electrical engineering in that it provides the basis for the modern generator, the telegraph, the telephone, the radio, and a host of other applications.

Thus in addition to inventing the electric motor, Faraday also discovered the fundamental laws that have made electric generators and most of the rest of electrical engineering possible. The cultural effects of his work upon the civilization of the twentieth century can scarcely be overestimated; for which reason he is often justly called the "father of electrical science."

The True Nature of Magnetism

At the time when Faraday was conducting his epoch-making experiments with magnets and electric currents, he was unaware of the true nature of either magnetism or electricity. Since his time we have learned that an electric current is simply a movement of electrons through a wire. The behavior of the electron also offers at least a partial explanation for the phenomenon of magnetism. We have learned that electrons moving in a wire produce a magnetic force around the outside of the wire. This magnetic force, therefore, must be something that is created by an electron in motion. Since the atoms of all material things contain electrons moving at terrific speeds within the atoms, why are not all substances magnetic? Why are not all things surrounded by a magnetic field? We know that many things, such as glass, wood, copper, and a host of others, have no magnetic properties; in fact very few materials do show such characteristics.

The answer to this apparent paradox goes back to a consideration of magnetic fields produced by moving electrons whose direction of motion may be accurately determined. Let us recall

The 100-pound girl is supported by the tiny magnet at the top of her swing. This special permanent magnet is a new alloy of aluminum, nickel, iron, and cobalt invented by General Electric scientists, and it will lift 500 times its own weight. (Courtesy of General Electric.)

the second part of Oersted's experiment. Electrons from a battery were flowing in a given direction through a wire. When he reversed the connections on the battery, thus reversing the direction of electron flow in the wire, the compass needle reversed its direction and indicated a reversal of the magnetic field, as shown in the drawings on page 419. Stated in modern terms, this means that the direction of the magnetic force depends upon the direction of motion of the electrons.

Should two electrons move past each other in opposite directions, the magnetic force of one would neutralize the magnetic force of the other. This is what generally happens within the atoms of most materials. However, in the case of a

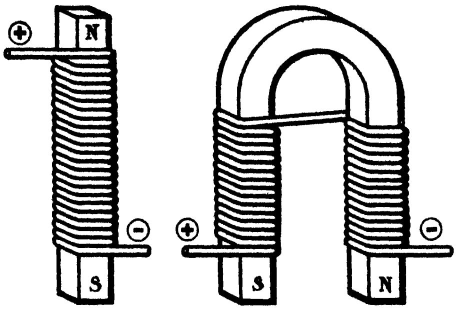
The latest type of mariner's compass. (Wide World photograph.)

few substances most of the electrons within the atom apparently move in the same direction. Such uniformity of electronic behavior within the atom is unusual, but it does take place in a few such substances as the atoms of iron, cobalt, and nickel. Under such conditions, each atom of the substance is surrounded with a magnetic force of its own, and it behaves like a tiny magnet. Should enough of these atoms line up in one direction, their individual magnetic forces would act in unison, and the entire substance would then behave as a single magnet. In some cases individual atomic magnets retain their alignment for a very long time; they are called "permanent" magnets.

Any permanent magnet is characterized by the tendency for one end to point toward the north and the other toward the south. For this reason one end of a magnet is called a north pole, and the other a south pole. When two magnets are brought together, the like poles repel each other, and unlike poles attract. This leads to an explanation of why a magnetic compass will indicate geographical directions. The earth is one magnet, and

the compass needle another, and they react on each other so that the compass needle turns in a north and south direction.

It is important not to confuse magnetic fields and electric fields. They follow similar laws of attraction and repulsion, but they are totally different. Electric fields are a permanent property of electric particles, whereas magnetic fields are a property of moving electric particles only.



Two kinds of electromagnet—bar magnet and horse-shoe magnet.

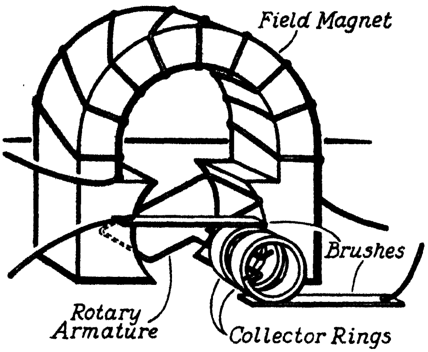
For practical use, powerful magnets may be made by wrapping a coil of wire around a bar of soft iron, as illustrated in the accompanying drawing. When a current is passed through the coil, the iron bar becomes strongly magnetized, one end being a north pole and the other a south pole. When the current is reversed, the position of the north and the south poles is reversed; and when the current is turned off, the magnetic force disappears. This type of magnet is called an electromagnet. Electromagnets are sometimes made in a U shape so that the north and south poles are adjacent, an arrangement that produces a strong magnetic field between the two poles.

The electromagnet has important uses in all types of electrical apparatus, since it provides a means of transmitting mechanical motion over electric circuits. For example, it is an electromagnet that converts a telegraph message into the audible "click" of the sounder or operates the teletype printer. Electromagnets ring the bell in a telephone and also convert the electric currents in the receiver into audible sound waves. Furthermore, the magnetic force essential to the operation of all generators and electric motors is produced by electrons flowing in coils of wire in the form of electromagnets.

The Modern Electric Generator

The purpose of the electric generator is to convert mechanical into electric energy. This can be done by using mechanical

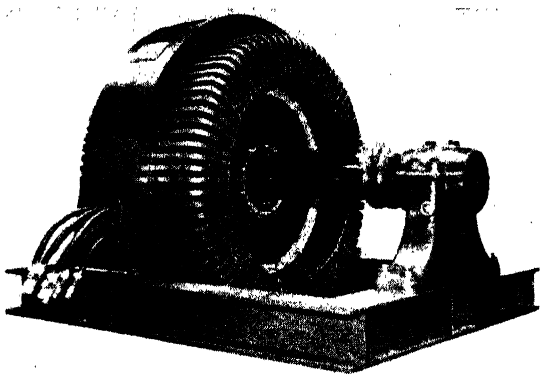
energy to move a conductor through a magnetic field, thereby inducing a current in the conductor. The accompanying drawing illustrates the design of a simple generator consisting of a loop of wire mounted on a shaft so that it can be rotated between the poles of an electromagnet. The rotating loop is called an armature and the stationary electromagnet is called the field. As the shaft is rotated by some source of mechanical energy, the relative motion between field and armature induces a current in the latter.



A simple alternating-current generator.

In order to utilize the current induced in the armature, it must be connected with an external circuit. Ordinary wires cannot be used for this connection, because they would be twisted off as the armature spins around. Accordingly, it is necessary to use a set of sliding contacts, or brushes, which make contact with the surface of a pair of rings mounted on the armature shaft, as shown in the drawing. Each ring is connected to one end of the armature loop so that the induced current can flow into the external circuit through one ring and return to the armature through the other. These rings are called collector rings, because they "collect" the induced current and pass it on to the brushes. In the illustration the armature is shown as a single loop or wire for the sake of simplicity. In an actual generator, the armature consists of several such loops all connected to the same pair of collector rings.

All the electric energy used for lighting and for driving electric motors is produced by generators of this general type. Nowadays all city dwellers as well as the residents of a great many rural districts accept electricity as a household necessity. If the electric service is interrupted for only a few minutes, both industrial and domestic activity are paralyzed. This fact is the best testimony to the importance of the generator in modern civilization. Without it the widespread use of electricity would be impossible, because batteries are much too expensive to be



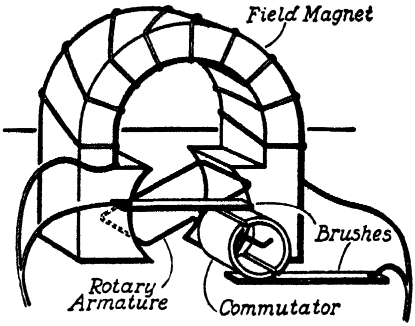
Electrical connection between the rotating armature and an external circuit is made by means of the collector rings and sliding contacts shown near the end of the shaft. (Courtesy of General Electric.)

used as a source of supply for a large amount of electric energy. The chief use of batteries at present is in portable equipment or equipment that must operate where electric power lines are not available.

More than half the electric energy produced by generators is utilized by various types of electric motors. The importance of the generator, therefore, is to a large extent dependent upon the fact that modern electric motors are available. Generators and motors are sometimes quite similar in appearance and construction, although they perform opposite functions. Both are equipped with a stationary field magnet and a rotating armature. The function of the field magnet is identical in both cases, but in a motor the armature acts as an electromagnet rather than as a carrier of induced currents.

The operating principle of any electric motor depends upon the fact that magnets exert forces upon one another. In practical motor design, it is necessary to devise a means of converting these forces into continuous motion. The essential elements of a modern electric motor are shown in the drawing of a simple motor. When current is supplied to the field magnet and the armature simultaneously, two separate magnetic fields are produced. These fields are made to interact with opposing forces and thus cause the armature to rotate so that its north and south poles take the position shown in the diagram. To prevent it

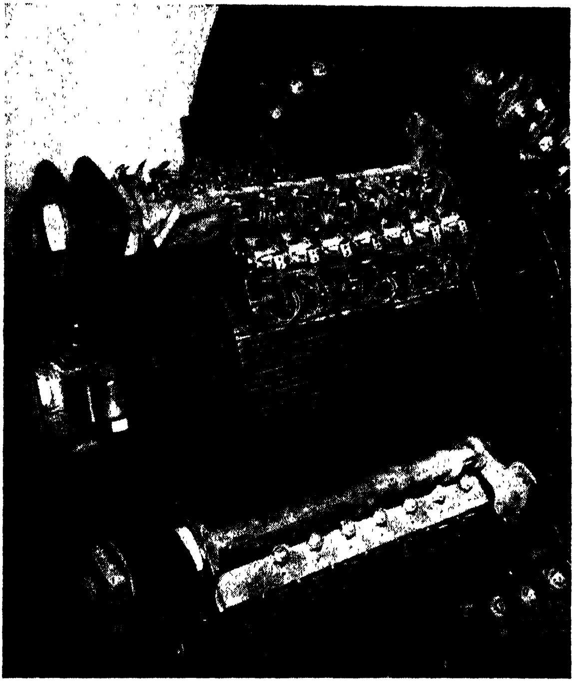
from stopping at this position the flow of current in the armature is reversed at the instant when the armature gets to this position. This in turn reverses the magnetic poles of the armature, and the armature is forced to seek a new stopping place one-half turn farther on. Reversal of the current in the armature is again repeated, and thus the motion is made continuous.



A simple electric motor.

The current reversal is accomplished automatically by a special type of switch mounted on the armature shaft, as shown in the drawing. This is called a commutator. Note that it consists of two semicircular segments insulated from each other, one segment being connected to each end of the armature coil. As the armature rotates, each of these segments alternately moves from one brush to the other at just the right instant and thereby automatically reverses the flow of current in the armature twice during each revolution. Thus the armature never finds a stopping place, and its rotation continues. For simplicity, only one armature coil has been shown in our diagram. In an actual motor the armature consists of several coils, and the commutator contains two segments for each coil. The segments are made correspondingly shorter so that they can all be fitted into one cylindrical structure.

Modern electric motors are available in sizes ranging from the tiny ones used in electric clocks up to several hundred horsepower. They have reduced the use of muscular effort to a minimum, because motors can be designed to do any kind of mechanical work. Another important feature is the fact that it is easy to supply energy to them. Electric lines can be installed anywhere in a building from the basement to the roof; and once the installation is made, the power supply is permanent. Large quantities of power can be generated economically at a central powerhouse and transmitted over wires to thousands of individual users for consumption. It is the electric motor that is

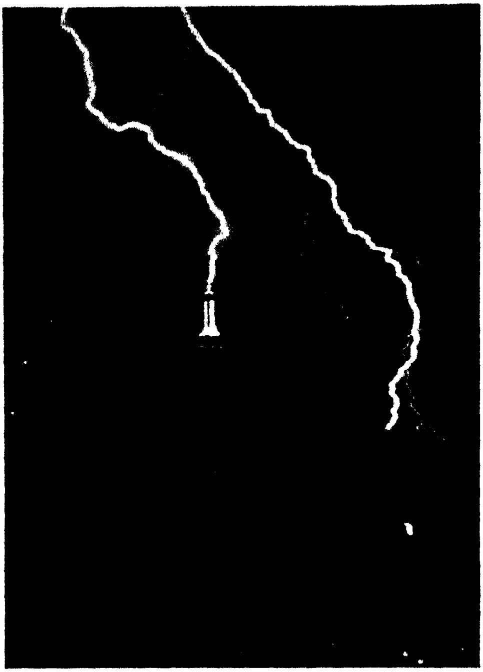


Brushes and commutator of many segments on a large direct-current motor. Each pair of segments in the commutator is insulated from all others and is connected to a separate coil in the armature. (Courtesy of General Electric.)

primarily responsible for making the twentieth century an "electric-button" age.

Converting Electricity into Light

The discovery that electricity will produce light dates back several hundred years. It may have been first observed in the form of sparks escaping from a charged object. Benjamin Franklin conducted many familiar experiments with electric sparks. By flying a kite during a thunderstorm and drawing sparks from the end of the kite string he established the electrical nature of lightning. He observed that the sparks produced a flash of light, but he knew of no practical way of generating them fast enough to give a steady glow like the light from a candle. Not until about 1810 did Sir Humphrey Davy discover how to produce a steady discharge of electricity across a small



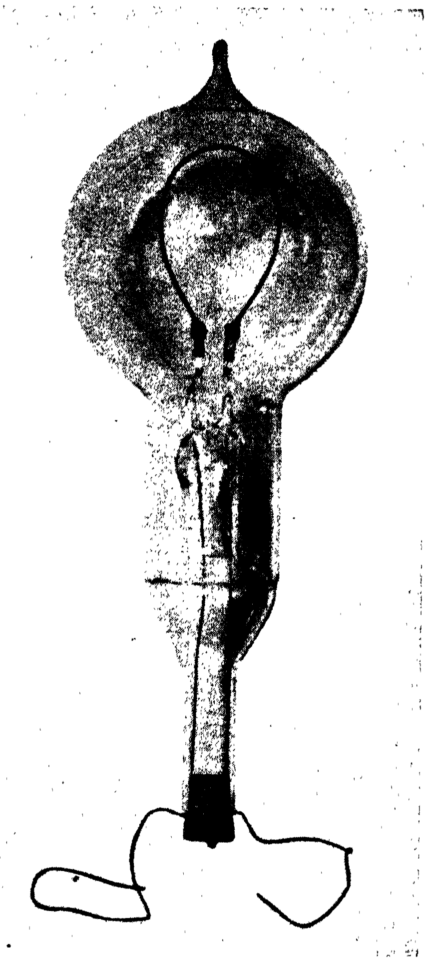
Lightning striking the Empire State Building during a severe thunderstorm in the spring of 1934. (Photograph by Dr. K. W. Ney.)

gap between two conductors. His device constituted an electric arc, and it produced a very brilliant light. Davy used it as a piece of demonstration apparatus; but since the only source of electricity available at that time was the crude and expensive galvanic battery, it probably did not occur to him that electric arcs could be used as a practical source of illumination.

Several decades after Davy's invention, when suitable generators had been perfected, attention was once more turned to the electric arc as a means for converting electricity into light. One of the first practical systems was invented in 1878 by C. F. Brush of Cleveland, Ohio. The Brush system was applied quite extensively to the lighting of streets, public halls, and factories. Although electric arcs are a brilliant source of illumination, they have several drawbacks. The burning electrodes are exposed to the open air so that they are a potential fire hazard; also, considerable accessory apparatus is required to make their operation continuous, and even then there is some sputtering

and flickering. They were considered the last word in electric lighting in 1880; but because of their inherent limitations, they were not destined to hold this position long.

The name of Thomas A. Edison is associated with many of the electrical devices that we use today, but he probably is best known as the inventor of the incandescent electric lamp. Convinced from the start that the arc lamp was not the ultimate answer to electric lighting, he expressed his dissatisfaction by referring to it as a "commercial short circuit." His ambition was to build an electric light "within a bottle," an ambition that was realized when he brought out his first carbon-filament incandescent lamp about 1880. The Edison lamp consisted of a fine filament inside a glass envelope from which the air had been exhausted. The filament had a high resistance which caused it to get hot enough to glow with light when an electric current flowed through it. Carbon-filament incandescent lamps were crude and inefficient compared to the electric lamps we are using today, but they were the beginning of a development in electric lighting that spread to the far corners of the earth and rendered arc lamps obsolete, except for special applications. The principle used by Edison is still employed in all incandescent lamps. The improve-



Operating replica of Edison's early incandescent lamp, showing carbon filament in glass envelope and connecting wires. (Courtesy of General Electric.)

ments have been primarily in getting better filaments which could be heated hotter so as to give a whiter light.

The incandescent lamp has been brought to its present state of perfection by painstaking research which has been conducted in the leading research laboratories of the world for a period of half a century. Nowadays there is a lamp for every type of application, from the tiny "grain-of-wheat" lamp used on surgical instruments to the giant beacons that guide ships and airplanes to safe landings in the darkest nights. The fact that four-fifths of a billion incandescent lamps were sold in the United States during 1939 is testimony to their widespread use not only as an aid to vision but for all sorts of decorative effects in the fields of entertainment and advertising.

Despite its position of supremacy in the field of lighting, the incandescent lamp is a relatively inefficient device. Most of the electric energy supplied to it is converted into useless heat. This waste is unavoidable, since the efficiency cannot be improved without increasing the filament temperature to a point where the filament would melt away in a few minutes of operation. Therefore, in order to produce a more efficient electric lamp some new principle must be employed. Within the last two years this has been accomplished with the introduction of the "fluorescent" type of lamp.

In the discussion in Chap. 12 it was noted that invisible ultraviolet rays cause certain minerals to "fluoresce," or glow, with a visible light. In the fluorescent lamp, invisible ultraviolet radiation is produced in a long glass tube by means of a mercury arc. The inside surface of the tube is coated with fluorescent minerals which glow with a brilliant fluorescent light when the ultraviolet rays strike them. By proper choice of fluorescent materials, these lamps can be made to produce white light, daylight, or any desired color effect by the additive process. They are inherently more efficient than incandescent lamps, the efficiency being about twice as great for white light and still greater for any single color. Since these lamps also generate much less heat than the incandescent type, they can be mounted in closer quarters and with less provision for ventilation. Their low surface brightness reduces glare, and their flexibility in color production makes them ideal for display and decorative lighting.

In view of these advantages it is possible that the introduction of the fluorescent lamp in 1939 marks the beginning of a new era in artificial lighting.

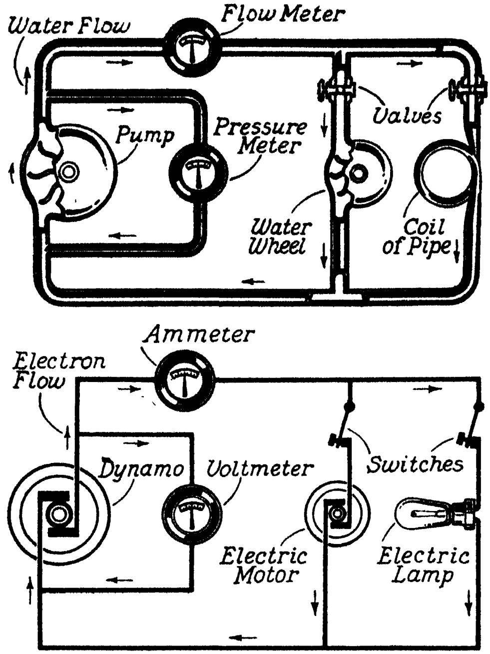
Electrical Language

Like any other well-developed branch of science, electricity has a terminology all its own. It is not the purpose of this brief survey to acquaint the reader with the complete vocabulary of the electrical engineer, but some important things about electricity are difficult to understand without a knowledge of a few specific terms.

Everyone knows that electric current flows in a wire, but how many are aware of the fact that in order for it to get anywhere it must flow in two wires at the same time? Nevertheless, if you examine every kind of electrical device, from a floor lamp to a trolley car, you will see that it is supplied with current not by one wire but by a pair of wires. The reason for the necessity of two wires goes back to the fact that electrons must flow from negative to positive terminals. In using electricity we must make them flow from the negative terminal of the generator through a wire into the device that we are using, which may be an electric lamp, a motor, a radio, or an electric iron. The electrons do not "pile up" in the device; instead, they simply flow through it and must flow back to the positive terminal of the generator through another wire. Therefore, one wire brings electrons to the electric appliance, and the other carries them away.

Incidentally, these wires must be thoroughly insulated from each other with rubber or similar insulating material. If insulation is not present and the wires come in contact with each other, we have what is commonly known as a short circuit, a condition accompanied not only by fireworks but sometimes by serious shocks, burns, and fires.

It is easier to understand why all this is true if we compare the flow of electrons to the flow of a fluid like water. In the top drawing of the accompanying illustration we have shown a pump arranged to circulate water through a system of pipes. The general operation of the system is quite obvious. As long as mechanical energy is supplied to the pump, it will force water into the upper horizontal pipe. The pressure meter will indicate how hard



Electron flow in a wire circuit may be compared to the flow of water in pipes.

the pump is pushing against the water, but there will be no actual flow of water unless one of the valves is opened. If the left valve is opened, water will be forced down through the water wheel, and its energy will be used up in turning the wheel. The water itself returns to the pump through the lower pipe. If the right valve is opened, water is forced down through the coil of small pipe. Since the pipe is small, the water has a hard time getting through it, and the energy of the water will be converted into heat because of friction against the walls of the pipe. Again it should be noted that the water itself is returned to the pump. The rate at which water is flowing will be indicated by the "flow meter" mounted on the top pipe.

In the lower drawing is a type of sketch that is technically called an electric circuit diagram. The interconnecting lines represent wires, and all the other parts are properly labeled. The behavior of this circuit is exactly the same as the water system just described; the only difference is in the words used to describe it. When mechanical energy is supplied to the generator, it

pumps electrons away from the positive terminal and piles them up on the negative. The electronic pressure thus developed is measured in terms of "volts," as indicated by the voltmeter connected between the two terminals of the generator. The electrons cannot flow anywhere, however, as long as both of the switches are turned off.

When the left switch is closed, imprisoned electrons will rush through the electric motor, and their energy will be converted into mechanical energy. The electrons themselves, however, return to the positive terminal of the generator. Should the right switch be closed also, electrons would be forced down through the lamp filament. They encounter considerable resistance in going through the extremely fine filament, and their energy is converted into heat and light of the glowing filament. Again it should be noted that the electrons themselves are not used up; rather, a part of their energy of motion is converted into the light that we use. The electrons return to the generator to be pumped around again.

The rate of flow of the current is a measure of the number of electrons per second passing through the circuit. It is measured in terms of a unit called the "ampere," equal to about six and one-fourth billion billion electrons per second. The current in amperes is indicated on the ammeter shown in the upper wire of the circuit diagram. Now, it should be evident that the amount of electric current that will flow around a circuit will depend upon the resistance encountered by the moving electrons. This is such an important factor in electrical engineering that a special unit is used to measure resistance; it is the "ohm."

These three units—volts, amperes, ohms—measure three essential elements in every flow of electricity through an electrical circuit. The volts measure the electric pressure; the amperes, the size of the current flowing; and the ohms, the resistance encountered. They are of such fundamental concern as to constitute the ABC's of every electrical engineer's vocabulary. Moreover, for any given circuit, these three quantities are related in a definite way, the relationship constituting one of the most fundamental laws of electricity. It was discovered in 1827 by a notable German scientist, George Simon Ohm, and is known as Ohm's law. It states that the current flowing through any

circuit is equal to the volts applied to the circuit divided by the ohms of resistance in the circuit, and it may be written in an abbreviated form as

$$\text{Amperes} = \frac{\text{volts}}{\text{ohms}}$$

Thus it is easy to see that there are two ways of increasing the flow of current in any electric circuit. It can be done either by increasing the voltage that is applied to the circuit or by decreasing the resistance of the circuit. Did you ever see a defective electrical appliance suddenly throw off smoke and sparks or perhaps even catch fire? The incident may have been explained to you by the terse expression, "There is a short circuit somewhere." This is correct, but what it actually means is that the resistance between the two wires suddenly became very low: *i.e.*, the path of the current was "short," and the sudden increase in current resulting from the reduced resistance generated enough heat to produce the accompanying fireworks. Short circuits are usually caused by defective insulation at the point where the cord enters the appliance or where it enters the wall plug, but they may occur anywhere. To prevent any great damage from such an occurrence, all properly installed wiring is protected by fuses that melt and open the circuit the instant that the flow of current gets dangerously high so that no more current can flow.

How Much Does Electricity Cost?

Since the electrons constituting an electric current always return to the generator, it might be reasoned that electricity is never sold. In fact, a student once asked why the power company insists on charging the consumer for electricity as it always gets all its electrons back again. The answer to this question is that electricity is sold as a service, and the sale does not involve the transfer of a physical commodity, as is the case with a sale of coal, flour, or sugar. This service consists essentially in supplying energy to the electrons so as to drive them through lamps, motors, and similar appliances. The power company provides the generators, power lines, and other facilities necessary for this service, and furnishes the consumer with as much

electric energy as he may require. Since the cost of the service is charged on the basis of the amount of electric energy consumed, it is necessary to have some satisfactory means for measuring this energy. Neither a measure of amperes nor a measure of volts is sufficient for this purpose; for both of these are involved in electric energy.

For a given circuit, the rate at which energy is consumed is equal to the product of the applied voltage times the current flowing through it. It is measured in a unit called the "watt." The simple formula for watts is

$$\text{Watts} = \text{volts} \times \text{amperes}$$

For example, a 100-watt lamp consumes electric energy four times as fast as does a 25-watt lamp. Please note that we did not say that the 100-watt lamp consumes four times as much electric energy as the 25-watt lamp does but rather that it consumes energy four times as fast. The larger one would consume four times as much only when both were left burning for the same length of time. The total amount of electric energy consumed by any device is equal to the rate of consumption multiplied by the time of operation. Translated into electrical language, this can be expressed as follows:

$$\text{Total energy} = \text{watts} \times \text{hours}$$

The expression has been abbreviated in practice to the term watt-hours, and it is the basic unit for the measurement of electric energy, just as the calorie or B.t.u. is the basic unit for heat energy. The watt-hour represents a rather small quantity of electric energy. For example, if you have a sixty-watt bulb in your study lamp, it uses one watt-hour every minute. For convenience, therefore, the power companies measure electricity in 1,000-watt-hour chunks, a unit that is known as the kilowatt-hour.

Every subscriber to electric service is provided with an electric meter that records the total number of kilowatt-hours consumed in the subscriber's circuit. The mechanism for making this record is similar to the mileage recorder on an automobile speedometer. At a given time each month the meter is read, and a bill is rendered for the difference between this reading and the reading for the previous month.

Kilowatt-hour meters, one at right in operation. (Courtesy of General Electric.)

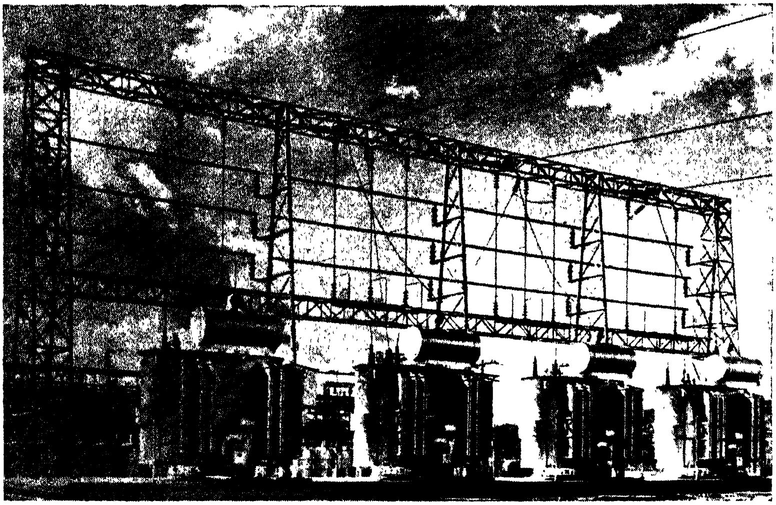
Electric Energy Lost in Transit

Practically all electric household appliances in the United States are designed to operate on 120 volts; accordingly, the power company supplies electric energy at this voltage to the consumer. Since the voltage remains relatively constant, the current flowing through any appliance may be found when the wattage is known. For example, the reader may find it an interesting, simple mental process to verify the fact that a 60-watt lamp has a current of one-half an ampere flowing through it, whereas a 600-watt electric flatiron draws a current of five amperes. It was stated earlier in the chapter that an electric current tends to produce heat in any wire through which it is flowing. Some of this electric energy is dissipated

by simply causing heat in the connecting wires, but the loss due to the heating effect is seldom more than about one per cent, provided the wires are large enough. For this reason, all appliances are fitted with connecting cords that are amply large for the current that they are to carry. You may have noticed that the cord on an electric iron, for example, is considerably heavier than that on a small floor lamp.

The problem of adequate wire size is never serious in the average household, because the distances to be covered are small, and the total number of watts used by any one appliance is never more than a few hundred. However, when large quantities of electric power are to be transmitted from a power station to buildings miles away, the story is altogether different. The simplest way to illustrate this difference is with a practical example. Suppose that a power station has a group of generators with a total output of 60,000,000 watts, a condition not unusual for a large power company. It is also not uncommon to transmit this energy for at least a mile. Some loss of energy in the transmission line is unavoidable, but it is desirable not to allow it to exceed about 4 per cent. The best commercial material for efficient and economical transmission of electricity is copper; so copper wire is used almost universally. How large do you suppose that the copper wires would have to be to transmit 60,000,000 watts one mile at 120 volts, assuming a 4 per cent loss in transit? This would mean designing a pair of wires large enough to carry a current of 500,000 amperes for one mile with only 4 per cent heating loss. We shall not tire you with the details of the calculations, but the answer may be of interest. Each wire would have to be about nine feet in diameter! The total weight of copper in the pair of wires would be 170,000 tons and would probably cost around $100,000,000!

Now, one does not have to be an electrical engineer to know that these figures are unreasonable. If this were the only practical solution to long-distance electric power transmission, there would not be any. Each building would have to maintain its own small power plant, and the distance over which electricity could be economically distributed would be measured in feet rather than in miles. On the other hand, we do know that enormous quantities of electric power are transmitted efficiently



Substation transformers for handling 25,000 kilovolt-amperes. (Courtesy of General Electric.)

for distances of hundreds of miles. What is the answer to this apparent paradox?

It can be summed up in the familiar phrase, "high voltage." To be a little more specific, large quantities of electric power cannot be transmitted efficiently over long distances at voltages as low as 120 volts. The amount of copper required in a transmission line increases as the square of the current flowing through it. The practical solution to long-distance power transmission, therefore, is to keep the voltage on the transmission line as high as possible, so that the current will be small. For example, if 60,000,000 watts is transmitted at a voltage of 120,000 instead of 120, the current required will be only 500 instead of 500,000 amperes. Only about one-millionth as much copper would be required to carry this current; and although the cost would not be reduced in direct proportion, it would be brought down to a reasonable figure.

Transforming Voltages

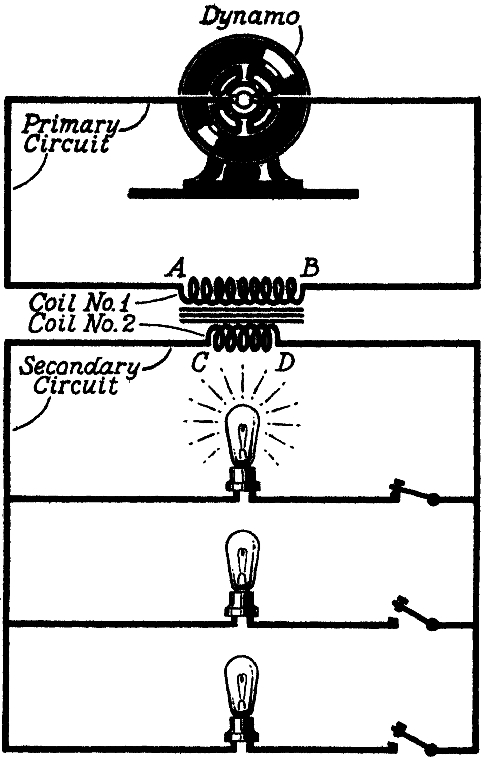
Putting this solution into practice is not so easy as it may sound. Even if we assume that electricity can be generated at extremely high voltages, for reasons of both economy and safety

it cannot be supplied to the consumer at much more than 120 volts. Appliances designed for high voltage would be expensive; moreover, unless handled very carefully, each appliance would be a potential electric chair. Some kind of compromise is necessary. In practice this consists of separating the long transmission line from the shorter subscriber's line so that each may be operated at the optimum voltage.

There are various ways of doing this, but the methods generally employed involve the use of electricity in the form known as alternating current, usually abbreviated as "a.c." In our discussion of electric circuits earlier in this chapter we implied that electrons flow continuously in one direction only. This is always true provided the generator has fixed positive and negative terminals. Such generators are said to produce direct current, usually designated "d.c." This is current that always flows in the same direction around the circuit. Most modern generators, however, are arranged so that the two terminals change polarity, *i.e.*, become alternately positive and negative, one hundred twenty times a second. This means that the electrons do not flow in one direction continuously but alternate, or surge back and forth, in the circuit. Each electron goes through a complete alternation in direction of flow every one-sixtieth of a second, so that the current is said to have a frequency of sixty cycles per second.

The advantage of this type of operation is that electric energy so produced may be transferred from one circuit to another and the voltage changed while the transfer is being made. Just how this is done may be shown with the drawing on the next page. Here we have two separate circuits that are not connected in any way by direct contact. Each circuit contains a coil of wire, and these coils are placed close together. In practice they are actually wound, one on top of the other, on the same iron core. The wires of each coil are well insulated so that there is no electrical contact between them.

The generator supplies alternating current to coil 1. As the electrons surge back and forth through this coil, the magnetic field around it surges back and forth also. Now, according to Faraday's principle of induction, this moving magnetic field will induce or create a current in coil 2. Therefore, the electrons



The principle of the transformer.

will start surging back and forth in circuit 2 at the same frequency as in coil 1. The voltage applied between terminals *A* and *B* on coil 1 is equal to the voltage supplied by the generator. The voltage induced across coil 2, *i.e.*, between terminals *C* and *D*, depends upon the number of turns of wire on this coil as compared to the number of turns on coil 1. If the two coils have an equal number of turns, then the voltage induced in No. 2 will be equal to the voltage applied across No. 1. If No. 2 has only one-tenth as many turns, it will develop only one-tenth the voltage of No. 1. If, on the other hand, No. 2 has ten times as many turns, ten times the voltage will be developed between its terminals.

It is seen, therefore, that this double-coil device will change the voltage of alternating current. It is called a transformer, because it "transforms" voltage from one value to another.

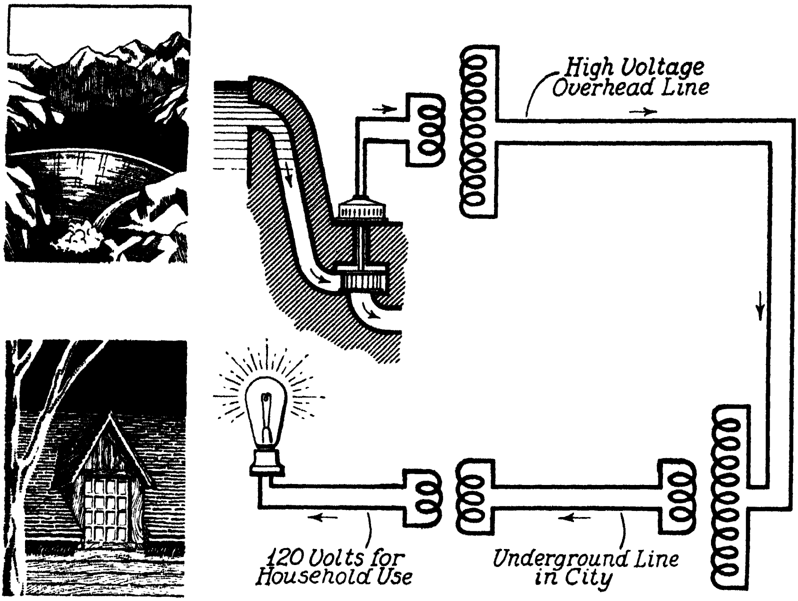
Coil 1 is referred to as the primary coil, and coil 2 as the secondary. A transformer in which the secondary coil has more turns than the primary is a step-up transformer, because it increases the voltage. One in which the secondary has fewer turns than the primary is a step-down transformer, because it reduces the voltage.

A transformer may be used to produce any desired change in voltage, but it does not change the electrical power of either primary or secondary circuits. Although the voltage across the secondary coil of a step-down transformer is low, this coil has a relatively small resistance, so the current that can flow through the secondary circuit is large. Thus the low voltage is compensated for by a large current, and the total watts are about the same as in the primary. In the case of the step-up transformer just the reverse condition is true. The secondary coil develops a high voltage; but since its many turns give it a high resistance, only a relatively small current flows through the secondary circuit. Thus the high voltage is offset by a low current, and the watts still remain about the same. It should be clear that a transformer does not generate electric energy; its function is merely to transform voltages from low to high or from high to low.

Now, referring once more to the drawing of the transformer, let us consider circuit 1 as representing the transmission line from the power plant and circuit 2 as representing the wiring in the subscriber's home. This will make it easy to understand how alternating current and the transformer have solved the problem of economical transmission of electric power over long distances. High voltages and low-ampere currents are sent out over the long lines. By means of a transformer the voltage is stepped down as the electric power is fed to the subscriber's line.

It should be quite obvious that the transformer will not work on direct current. Its operation depends on the constantly shifting magnetic field around the primary coil produced by an alternating current. In direct current this field is stationary, so it would be impossible for it to induce any current in the second coil.

The way in which alternating current and the transformer are applied to an actual installation is illustrated in a simplified



How the transformer makes it possible to transmit electric energy over long distances.

representation of a hydroelectric generator, power lines, and a consumer's home. Electric energy generated by water power at Niagara Falls is economically transmitted hundreds of miles to supply distant cities. Several transformers are employed so that all parts of the system can operate at the most suitable voltage. Without a means of changing voltage, the large-scale development of natural water-power sites would never have been possible. The vast amounts of energy available at such locations obviously cannot be consumed in the immediate neighborhood of the hydroelectric plant. So the existence of such enormous electrical powerhouses is made practical through this economical means of transmitting electricity over long distances to the cities where it is needed.

REFERENCES FOR MORE EXTENDED READING

Morgan, A. E.: "The Pageant of Electricity," D. Appleton-Century Company, Inc., New York, 1939.

The development of electricity from earliest times to the present told in an interesting story of the lives and achievements of the men who have contributed to the industry.

BRAGG, W. L.: "Electricity," The Macmillan Company, New York, 1936, Chaps. I–IV, inclusive.

An explanation of the elementary principles of electricity that is excellent for the beginner. Whenever a technical term is introduced, it is carefully explained so that its use will not be confusing.

LEMON, HARVEY BRACE: "From Galileo to Cosmic Rays," University of Chicago Press, Chicago, 1934, Chaps. XIX–XXVII, inclusive.

Here is an opportunity for the general student to get a thoroughly accurate explanation of the principles of electricity, written in a style that stimulates reading interest. Numerous little sketches, many with a humorous touch, are an important supplement to the text.

ERYING, CARL F.: "A Survey Course in Physics," Prentice-Hall, Inc., New York, 1936, Chaps. XV, XVI.

The author describes in an easily readable style some of the more widely applied principles of electricity and illustrates them with practical problems.

OSBORN, FREDERICK, A.: "Physics of the Home," McGraw-Hill Book Company, Inc., New York, 1929, Chaps. XXVII, XXXII.

Recommended for the beginner who is interested chiefly in learning how household appliances work.

BLACK, N. H.: "An Introductory Course in College Physics," The Macmillan Company, New York, 1935, Chaps. XVII–XXVI.

The discussion covers all the more important topics of electrical science and is suitable for college students. Chapters referred to are well illustrated with drawings and photographs.

UNDERHILL, CHARLES R.: "Electrons at Work," McGraw-Hill Book Company, Inc., New York, 1933, Chaps. I–VIII, inclusive.

The underlying principles of electricity are thoroughly discussed in a semipopular manner so that they may be easily understood.

SMITH, A. W.: "The Elements of Physics," McGraw-Hill Book Company, Inc., New York, 1938.

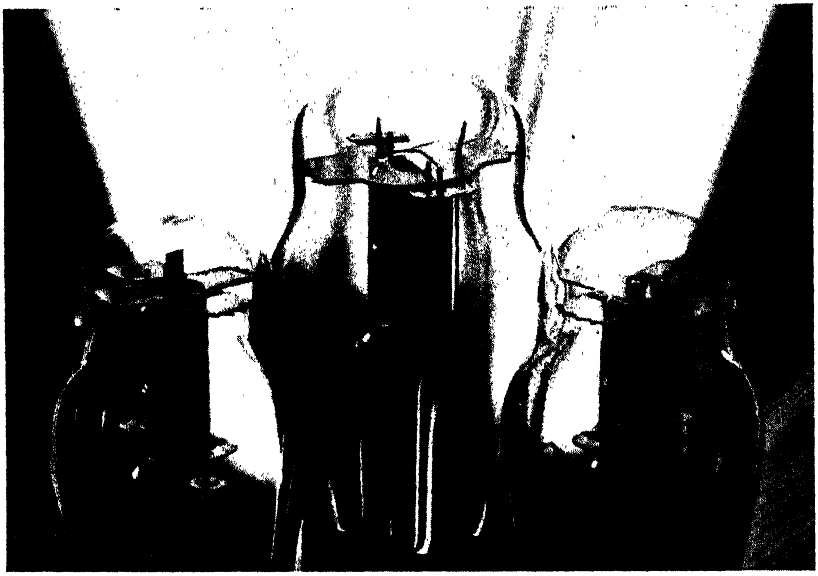
These chapters are perhaps too advanced for the beginner but are excellent for the student who has had an elementary course in electricity and wishes to pursue the subject further.

Electrical World, News Issue, published by McGraw-Hill Publishing Company, Inc., New York.

Concisely written articles expressed in nontechnical language and abundantly illustrated with photographs.

Technology Review, published by Massachusetts Institute of Technology, Cambridge, Mass.

One of the best journals in the field for articles relating to engineering development, most of them written in a style suitable for the intelligent student.



R. C. A. Laboratories.

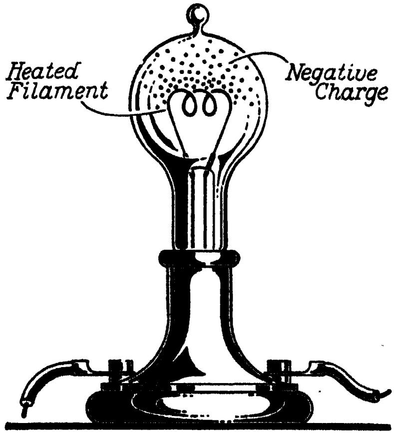
14: ELECTRONS IN GLASS HOUSES

Or the New Science of Electron Tubes

HAVE you ever examined the interior of your radio set? If so, you have observed that it contains glass tubes that look something like electric light bulbs. In fact when the set is turned on, these tubes also behave something like light bulbs in that they give off a faint reddish light. If you examine a lighted tube closely, you will see that the light is coming from a glowing wire near its center. Every radio user knows that if a tube does not light up in this fashion the radio will not operate until the burned-out tube is replaced with a new one. We can conclude, therefore, that an important relation exists between the light in the tubes and the operation of the radio. This is true, and the history of this relation goes back to Thomas A. Edison and his first incandescent lamp.

Electrons Escape from Heated Filaments

Edison realized that if his lamp was going to be a success, it must not only give light but must also last for a reasonable length of time. The carbon-coated filament would last long enough, but there was a tendency for particles of carbon to leave the filament and deposit on the inside of the glass bulb. This blackened the bulb to such an extent that after a short period of use the light from the filament could not get through the glass.

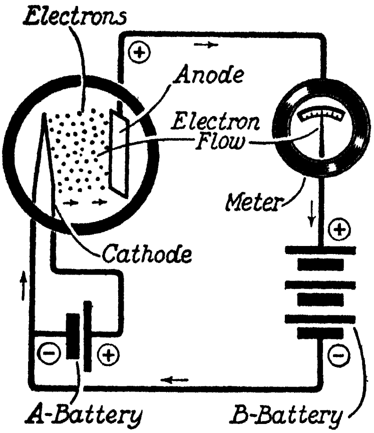


What Edison observed in his first electric lamps.

In the process of correcting this trouble, Edison made a number of experiments, one of which consisted of mounting a shield in the form of a metal plate between the filament and the inside wall of the glass bulb. When the lamp was lighted, he observed that this plate always became negatively charged, although it was not connected to any part of the electric circuit. Some kind of negatively charged particle apparently was coming from the surface of the filament to the plate. This was the first time that anything of this nature had been observed, and it became known as the "Edison effect." Subsequent experiments by Edison and many others proved that these escaping particles were electrons. It was found that electrons always escape from heated substances, particularly from certain metals and metallic oxides. Those thus set free are known as "thermionic" electrons, and the Edison effect is now known also as the thermionic effect.

By this time you may have surmised that the heated wire in a radio tube produces free electrons just as the filament in Edison's lamp produced them. But the question might be asked, Why are free electrons valuable in the electrical and related industries? The answer is that free electrons in a vacuum tube can be controlled with much greater speed and precision than can electrons flowing in a wire. The operation of any electrical

device as complex as a radio receiver depends upon a precision of electron control that is afforded only by the free electrons produced in tubes of this type. This principle of operation is responsible for the tubes frequently being called "electron" tubes, and in most of them the free electrons are produced by the thermionic effect.



The diode consists of only two electrodes, which when properly connected in an electric circuit produce an electron flow from cathode to anode.

One-way Streets for Electrons

Many laymen imagine that the operation of vacuum tubes in a radio receiver and similar equipment is complicated beyond reasonable understanding. This is far from the truth. A brief study of the principles of operation of the vacuum tube will open up a general insight into this basic element of a great many conveniences and necessities of modern life. Let us, therefore, look into this important and in some respects very interesting type of electron device.

Perhaps the best way to visualize the operation of a vacuum tube is to represent the simplest form of an electron tube and its associated electric circuits diagrammatically as shown in the accompanying drawing. The large circle represents the glass bulb of the tube. Inside the tube is a filament for producing electrons and a plate for collecting them. The battery for supplying heating current to the filament is called the A battery. It is represented by a long line for the positive terminal and a short line for the negative terminal, this being the customary symbol for a battery in any circuit diagram. The plate is connected to the positive terminal of another battery called the B battery. This battery is represented by several alternate long and short lines to indicate that it has a higher voltage than the A battery. The negative terminal of the B battery connects with one end of the filament so that the voltage of the B battery is applied between the filament and the plate.

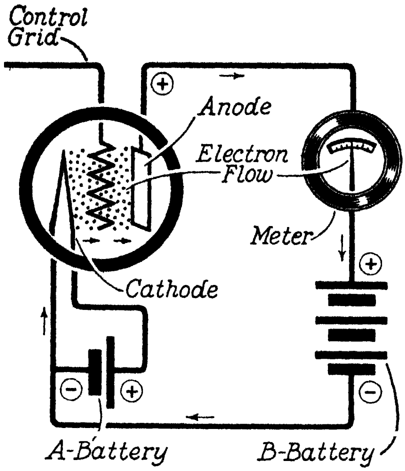
Now let us see what happens when the filament becomes heated. According to the thermionic principle, it produces a large quantity of free electrons. Since these carry negative charges, they are attracted over to the positively charged plate. They do not pile up on the plate; for if they did so, the positive charge on the plate would be neutralized. Instead, they flow down through the B battery and back to the filament whence they may again repeat the cycle. An electron circuit is established, and a current of electricity, which usually is referred to as the plate current, flows through the plate circuit. This current is indicated and its strength measured by the deflection of the electric meter in the plate, or B-battery, circuit. The filament in the vacuum tube is usually called the cathode, since it is connected to the negative side of the B battery, or power supply, and the positively charged plate is referred to as the anode.

Thus we see that a flow of electrons takes place from the cathode to the anode. Since electrons can escape only through the surface of the cathode, it is impossible for an electron flow to take place from anode to cathode, even if the B-battery connections are reversed. Therefore, this tube is a one-way conductor of electrons. A tube of this type is called a diode, because it contains only two electrodes. The property of one-way conduction makes it possible for the diode to convert an alternating current into a pulsating direct current. This is a necessary function for the detection of radio signals, as will be discussed in more detail in the following chapter. The diode also performs other useful functions in the field of electron control; moreover, its principle is of fundamental importance to an understanding of other types of electron tubes.

Adding a Third Electrode to the Diode

The number of electrons that flow from cathode to anode in a diode depends upon the voltage applied to the anode by the B battery, for a high voltage on the anode will attract more electrons than will a low one. Electron tubes have many uses, however, in which the flow of electrons must be varied without changing the voltage on the anode. In order that this may be done, it is necessary to add an extra electrode to the diode. This

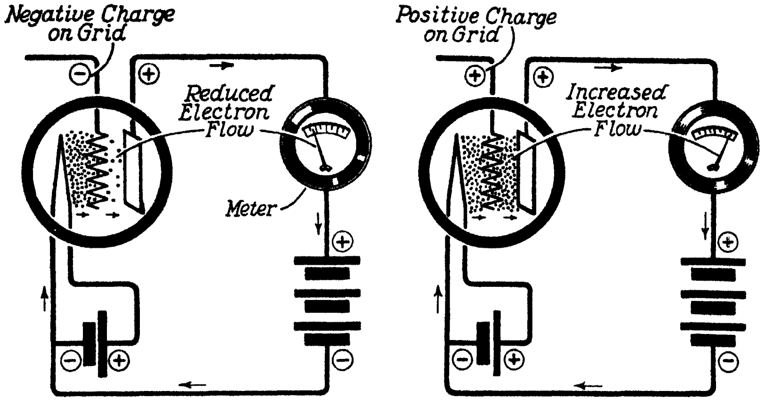
third electrode is called a control grid and is mounted between the cathode and anode as shown in the next diagram. Tubes of this general type are usually designated as triodes because they contain three electrodes. Structurally the grid is a meshwork of fine wires, so that it has relatively little solid surface. It acts as a sort of "traffic officer" to control the number of electrons flowing from cathode to anode.



The triode contains a third electrode known as the "control grid."

This is a unique service, yet the manner in which it is performed is based upon the simple and generally understood laws of electric attraction and repulsion. If a positive charge is applied to the grid, it exerts an attractive force on the electrons. They then speed across to the anode more rapidly and in greater numbers. Because of its open structure most of them flow right through the grid to the anode. On the other hand, if the grid is charged negatively, it exerts a repellent force on the electrons around the cathode so that only a few of them can get through and reach the anode. This controlling action of the grid is illustrated by the two diagrams in the drawing on the following page.

The important feature about a triode is the fact that a small voltage on the grid can control the same number of electrons as can a large voltage on the plate. This is because the grid is mounted much closer to the cathode, so that the force of its charge does not have to act through as great a distance as the charge on the anode. Thus the grid "amplifies" the effectiveness of a small charge in controlling electron flow. For example, a tube may be designed so that a variation of one volt on the grid will produce the same change in electron flow as a variation of ten volts on the anode. Such a tube is said to have an amplification, or "gain," of ten to one. In certain types of modern tubes the gain may be several hundred times.

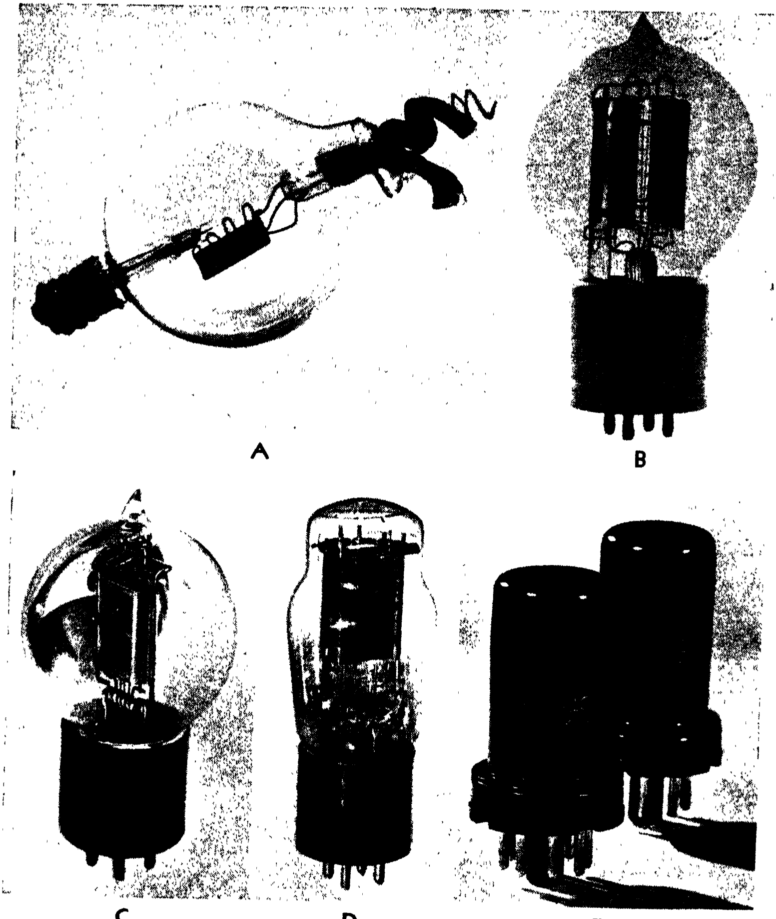


A negative charge on the control grid of a triode reduces the electron flow in the anode circuit, while a positive charge on the control grid increases the electron flow in the anode circuit.

The amplification in a vacuum tube of a weak electric charge is accomplished by applying the charge to the grid. We have just noted that this is much more effective than applying it to the anode. The size of the electron flow is directly proportional to the strength of charge applied to the grid; therefore, an exact duplicate of the weak-signal charge on the grid will be reproduced in the stronger current of the anode circuit. The fact that triodes can amplify or magnify electric energy does not mean that they create this additional energy out of nothing; instead, the amplified energy is supplied by the B batteries in the anode circuit. The triode is really merely an arrangement whereby very small amounts of electric energy applied to the grid can control much larger amounts from the B batteries going through the plate circuit. This gives the effect of a magnification of energy.

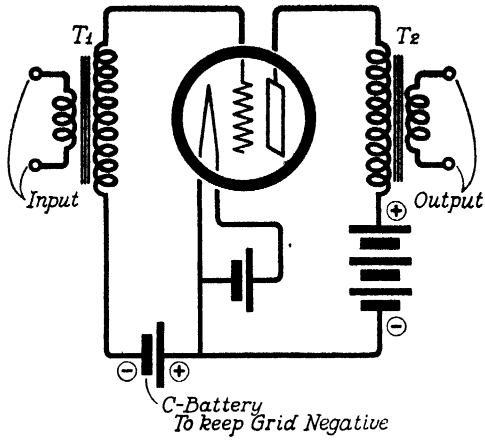
Most applications of electron tubes make use of their ability to amplify electric energy. Telephone, radio, sound movies, and television all involve the conversion of small amounts of electric energy into larger ones. The electron tube is the only apparatus that can accomplish this amplification satisfactorily, and this explains why vacuum tubes are so essential to the operation of these devices.

The manner in which a triode may be used to amplify a small alternating voltage or signal is illustrated in the accom-



Illustrating the development of vacuum tubes, (A) 1908, (B) 1918, (C) 1927, (D) 1938, (E) 1940. (Bell Telephone Laboratories photographs.)

panying drawing. The signal is fed to the grid by means of an input transformer (T_1). As the signal alternates back and forth, the grid becomes alternately positive and negative. The fluctuating grid charge causes the electron flow in the tube to increase and decrease so that a pulsating current flows through the primary coil of the output transformer (T_2) which is connected in the anode circuit. This pulsating current induces in the secondary coil a new alternating signal which is much stronger than the original signal applied to the grid of the tube; thus,



Single-stage amplifier with transformer coupling.

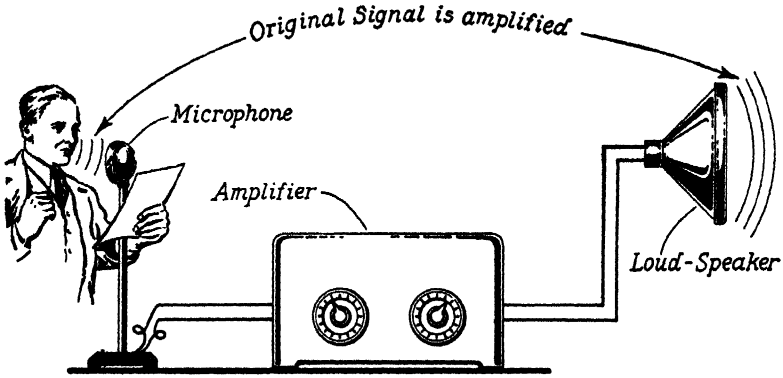
the signal has been amplified. By coupling several tubes together, one after the other, an amplification of several million times can easily be obtained.

The number of electrodes in an electron tube is not limited to three. In addition to the triode, we have tetrodes, pentodes, hexodes, and heptodes. These tubes contain additional grids so that the electron flow can be controlled to meet special requirements. However, the basic principle of amplifying a signal by means of a control grid is the same as just explained in the simpler tube and its accompanying circuit.

Reinforcing the Human Voice

Not so long ago the useful size of an auditorium was limited by the carrying power of the human voice. Other things being equal, the public speaker with the strongest lungs and the most endurance could command the largest audience. Today, thanks to the development of amplifying tubes, this is not true. It is possible to amplify anyone's voice to the point where it can be clearly heard in the largest of auditoriums or by open-air gatherings of thousands of people. Although this modern development may have accomplished little in the way of improving the content of public speeches, it has at least made speaking easier and listening more comfortable.

An apparatus for amplifying the human voice is called a public-address system. The three parts essential to its operation



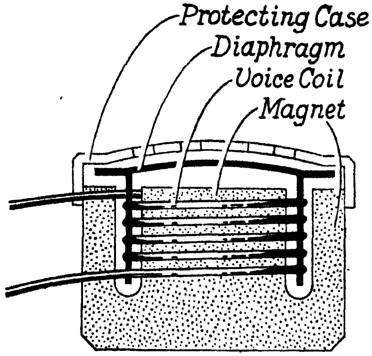
The essential elements of a public-address system are a microphone, amplifier, and loud-speaker.

are a microphone, an amplifier, and a loud-speaker, as represented in the accompanying drawing. Sound waves produced by the speaker's voice are picked up by the microphone and converted into small alternating currents. The currents are multiplied several thousand times by the vacuum-tube amplifier. The loud-speaker then converts these stronger currents back into sound waves that are much louder than the original ones entering the microphone. When properly installed, the amplification that can be produced by such a system is limited only by the cost of the apparatus. For large audiences, a powerful amplifier is used to operate several loud-speakers distributed over the listening area.

There are a great many different kinds of microphones, a wide variety of amplifier circuits, and several distinct types of loud-speakers that may be used in assembling a public-address system. To take up an explanation of all of them would lead us much too far afield; besides, understanding them is of primary concern only to the specialists in this field. It is in order, however, that we note here briefly the manner of operation of one kind of each of these devices. To do so may help to eliminate the mystery in some readers' minds as to how it is possible to convert the energy of sound into electric energy, amplify this energy, and then reconvert it into sound.

The microphone that we shall select for discussion is the dynamic microphone; but the name is not of significance except

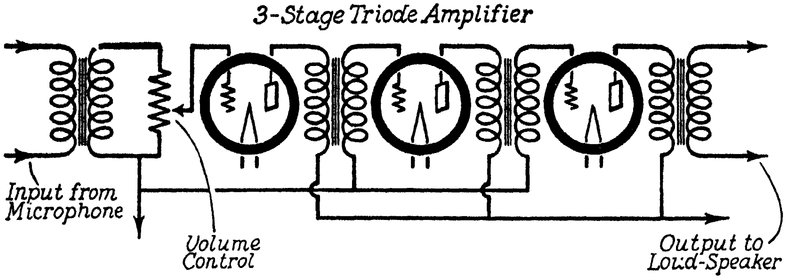
to identify it. Reference to the drawing may help visualize its construction and method of operation. This microphone consists of a coil of wire mounted on the back of a thin metal diaphragm. Around the coil and inside it are the pole pieces of a powerful permanent magnet. Sound waves falling on the diaphragm cause it to vibrate with the frequencies of the sound, and the attached coil moves back and forth in the field of the magnet with these same frequencies. This induces in the coil an alternating current that varies in frequency and intensity in exactly the same way that the sound varies. An alternating current, called the voice current, that has all the frequency and intensity characteristics of the sound is thus produced in the microphone.



The microphone converts sound waves into alternating electric signals.

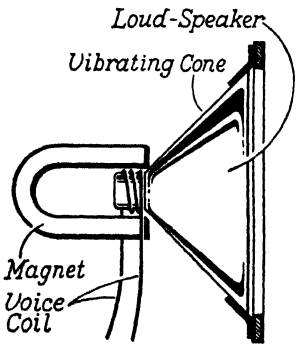
Wires from the coil carry this voice current to the input of the amplifier. We shall consider an amplifier of three triode tubes connected by means of transformers, as shown on the next page. This circuit illustrates well the principles of amplification and is not too difficult to understand. The voice current from the microphone flows through the first, or input, transformer and is impressed upon the grid of the first tube. This tube produces a corresponding but amplified current in its anode circuit which flows through the primary coil of the second transformer, where it induces a similar current in the secondary coil of this transformer. The current from this transformer is then applied to the grid of the second tube, and the amplifying procedure is repeated. Likewise, the same thing happens in the third tube. The current received by the output transformer is much stronger, therefore, than the original signal from the microphone. If each tube amplifies the signal one hundred times, for example, the total amplification will be one hundred cubed, or one million.

The amplified current from the output transformer is fed into the loud-speaker where it is converted into sound. The



The microphone signal is greatly increased by an amplifier.

manner in which this is accomplished may be explained with the simplest type of loud-speaker, represented in the next drawing. It consists of a coil mounted inside a permanent magnet and attached to a cone that is free to vibrate. The flow of current through the speaker coil produces an alternating magnetic field around it which interacts with the magnetic field of the permanent magnet. Since the magnet is stationary and the coil movable, the interaction between their fields causes the coil to vibrate with the same frequencies as the alternating current flowing through it. The vibration of the coil serves to drive the cone forward and backward with the same frequencies. The result is that the cone sets up waves in the surrounding air that correspond to the original vibrating sound waves, except that they are much more intense; and the original sound is thereby reproduced in great volume.



The loud-speaker converts the increased signal from the amplifier into sound.

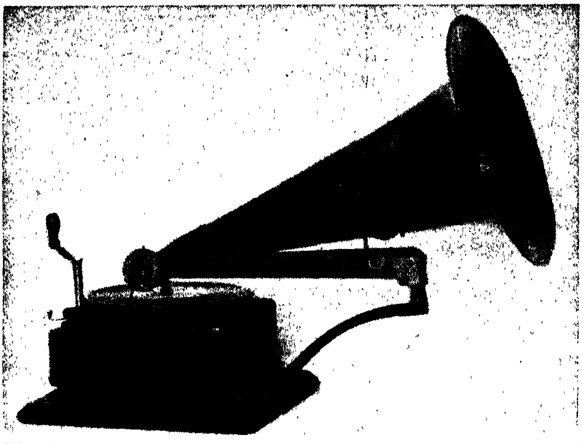
Public-address systems are a part of the equipment of any large auditorium, theater, or other place of assembly where speech or music is to be made audible to large audiences. When they are effectively installed and are the best that modern science can produce, they give a very faithful reproduction of the original sound, and the listener will not be conscious of their use except perhaps to realize that the sound is carrying better and farther than it could possibly do otherwise.

Public-address system installed in Saints Peter and Paul's Cathedral, San Francisco. Microphones were located in the pulpit and on the altar. Loud-speakers mounted on the pillars were designed to harmonize with the surroundings and provide a low-level amplification over the entire cathedral. (R.C.A. Laboratories photographs.)

Although public-address sets are sometimes similar to radio receivers in their outward appearance, these systems should not be confused with radio. The basic difference is that a public-address system does not depend on radio waves for its operation. Instead, it is actuated entirely by sound waves and alternating electric currents as described above. Since audible sound consists of frequencies ranging from about 30 to 20,000 vibrations per second, this range of frequencies is called the audio-frequency region. Public-address systems and similar equipment dealing only with the amplifications and reproduction of sound are designed to operate at audio-frequencies only and do not respond to radio signals.

Electrifying the Phonograph

In the early days of the phonograph the reproduction of sound from the record was perhaps entertaining, but certainly it was not particularly natural. Owing to limitations both in the

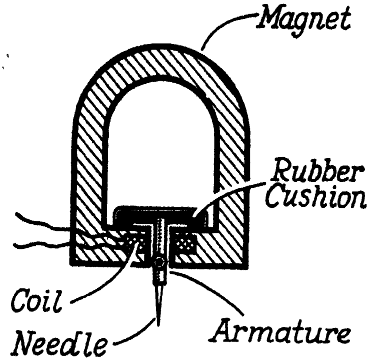


An early type of Victor acoustic phonograph. (R. C. A. Laboratories photograph.)

record and the reproducer, some of the frequencies present in the original sound were missing from the reproduced sound, *i.e.*, some of the original frequencies were lost in making the record, and additional ones were lost in playing back the record. Let us see what effect these missing frequencies have on the reproduced sound. If some of the high frequencies are eliminated, the sound loses the high-pitch notes which give it clarity and brilliance; if the low frequencies disappear, the bass notes which give sound its mellowness and much of its quality are missing. Each original sound contains a certain number of frequencies having a definite relationship to one another. Any disturbance of the relationship of the complete range of frequencies in the original sound produces distortion of the musical tones, and the old-fashioned phonograph had plenty of distortion.

Since the development of modern audio-frequency amplifiers, it has been possible both to record and to reproduce phonograph records with results that are so nearly like the original sound that only the best trained ear can distinguish the difference. Unfortunately, there are in use many electric phonographs that fall far below this standard, because good instruments are costly, and public demand is for something inexpensive. However, even the cheapest electric phonographs in use at present are a considerable improvement over the original machine invented by Edison.

The working principle of the electric phonograph is essentially the same as that of the public-address system, except that the microphone is replaced with a device called a phonograph pickup. The details of the electrical construction of one type of a pickup are shown in the accompanying diagram. Its principle of operation is essentially the same as that of the electric dynamo. The armature in this case is a small vane made of steel and mounted on the movable frame which supports the needle. The end of the vane opposite the needle is suspended between the poles of a permanent magnet. Surrounding this armature and also between the poles of the magnet is a small coil. Under these conditions any vibration of the armature causes a variation in the magnetic field strength between the poles, and induces alternating current in the coil. When the needle is placed in the groove of a rotating record, it follows the path of this groove very accurately. The sides of the spiral record groove contain tiny, wavelike indentations corresponding to the original sound waves used in making the recording. As the pickup needle moves along this groove, these indentations cause the needle to vibrate back and forth, with the same frequencies as those recorded in the groove, and the vibrations are transmitted to the armature. The result is an induced alternating current in the coil of wire around the armature, and the frequencies of this current will be the same as those recorded in the spiral groove. This alternating current is then amplified and reproduced by the same kind of amplifier and loud-speaker used in a public-address system.



In one type of phonograph pickups, electric currents are generated in the coil by the vibrating armature that is mounted inside it.

The chief reason why the electric phonograph has more natural reproduction than the old-style "acoustic" phonograph is that the latter depended entirely upon mechanical amplification. The vibrations of the needle were transmitted directly to a

A modern radio-phonograph combination. (R. C. A. Laboratories photograph.)

diaphragm which, in turn, was mounted at the small end of a long horn. The vibrating diaphragm reproduced the sound; however, its small size made the sound very low in intensity. An increase in loudness was obtained by the long horn, wherein the air column was set into resonant vibration by the diaphragm. This kind of reproducing system will amplify some frequencies more than others, because both horn and diaphragm will resonate with certain vibrations better than with others. Distortion of the sound is thereby produced. In a vacuum-tube amplifier, it is possible to amplify all frequencies uniformly so that the resulting movements of the loud-speaker cone correspond exactly to the original vibrations, except that the amplitude is much greater. The result is a quality of reproduction limited only by imperfections in the record itself.

The high quality of recorded speech and music made possible by various types of modern electric phonographs has found many applications. In the entertainment field, recordings may be used in places where the entertainers cannot appear in person, thereby bringing the best talent before much larger audiences. Recorded lectures and speeches serve as a new teaching device in the field of education. Likewise, these high-quality records are providing mechanical robots with voices so that they can answer predetermined questions and give instructions without the aid of a human operator.

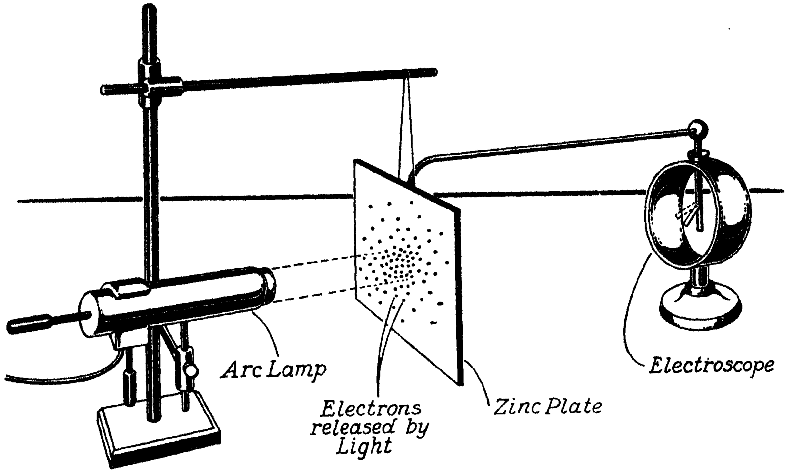
Birth of the Electric Eye

In 1887 Prof. Heinrich Hertz of Kiel, Germany, was experimenting with the transmission of electromagnetic waves. His experiments are now world famous, since he was the first person ever to produce wireless or radio waves in the laboratory, and they led to the discovery of another effect, unrelated to radio, which has been the basis for a modern development of great importance, namely, talking motion pictures.

The apparatus used by Hertz for detecting wireless waves consisted merely of a loop of wire containing a narrow spark gap. As the waves moved across this loop, a voltage was generated in the loop which caused an electric spark to jump across the gap. This part of the experiment was perfectly standard, and such a simple apparatus would seem unlikely to serve any purpose beyond the detection of wireless waves. In the hands of the average experimenter this probably would have been true; Hertz, however, was not an ordinary observer. Like all great experimenters he noted every minute detail about the behavior of his apparatus. On this occasion one of the points that puzzled him considerably was the fact that the spark across the gap was stronger when the apparatus was exposed to a bright light and weaker when it was enclosed in a dark box. Hertz realized that this effect of light on his apparatus had nothing to do with the wireless waves that he was investigating; nevertheless he considered it important enough to include a record of these observations in his notebook.

When Hertz published a paper on his studies of radio waves, he included a brief report of his observations of the effect of light on the discharge of an electric spark. Another German scientist, Hallwachs, read this report and decided to investigate the phenomenon further. He believed that there must exist between light and electricity a fundamental relation that was responsible for the effect observed by Hertz. He proceeded, therefore, to test the effect of light on the escape of electric charges from metals.

The apparatus used by Hallwachs is shown in the next drawing. The electroscope at the right side of the picture indicates the strength of an electric charge. When a strong charge is

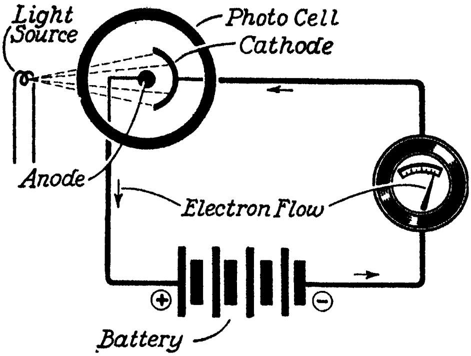


Light from the arc lamp removes electrons from the zinc plate and discharges the electroscope.

placed on the terminal at the top of the instrument, the thin metal leaf at the lower end of the terminal is repelled from the terminal rod so that it stands out in a horizontal position as indicated by the dotted line. If the charge is conducted away, the leaf gradually drops back to the rod. Hallwachs connected a metal plate to the electroscope so that the leaf would indicate any escape of electric charges from the plate. Provision was then made for exposing this plate to a beam of light from the arc lamp shown on the left side of the drawing.

Hallwachs' experiments consisted of charging the plate and observing the effect of light on the rate of discharge. When the plate was charged positively, the light apparently had little effect on the rate at which the charge disappeared. However, when a negative charge was placed on the plate, the exposure to light caused the leaf of the electroscope to drop rapidly. This indicated that the light rays made it easier for a negative charge to escape from the surface of the metal. The phenomenon was confirmed by repeated experiments and became known as the "Hallwachs effect."

About ten years later, the electron was discovered by Sir J. J. Thomson in England, and this discovery provided an explanation for the Hallwachs effect. When a metal absorbs light, some of the absorbed energy is imparted to electrons



A simple circuit, showing the action of a photoelectric cell.

in the atoms near its surface. This additional energy increases the motion of the electrons and makes it possible for a few of them to escape through the surface of the metal. Electrons released in this manner are called photoelectrons, and the Hallwachs effect has now been renamed the photoelectric effect.

The photoelectric principle is used at present in a type of electron tube known as the photoelectric cell, shown in the diagram above. It is similar in construction to the diode tube shown previously, except that the cathode consists of a plate of metal and the anode of a straight piece of wire. Electrons are released from the cathode by exposing the metal to a beam of light. Their release corresponds in principle to the action of the cathode in a radio tube except that in the latter the electrons are emitted when the cathode is heated, but in a photoelectric tube they are emitted where the cathode is exposed to light. A battery is connected between the cathode and anode of the photoelectric cell with the negative terminal connected with the cathode and the positive terminal connected with the anode so that when the electrons are released from the cathode, they flow across the tube to the positively charged anode. It is to be seen, therefore, that the photoelectric cell is an electron tube in which the electron flow from cathode to anode is controlled by a beam of light. The amount of electron flow, or strength of the anode current, is directly proportional to the intensity of the light falling on the cathode. Here, then, is a device for converting light energy into an electric current; furthermore, the strength of the current so produced is a measure of the intensity of the

light. Since it is possible to apply this device to many tasks that would normally require human vision, it is popularly called the "electric eye."

Many applications of the photoelectric cell involve its use in opening and closing electric switches. The current from the photoelectric cell, usually after amplification, flows through the electromagnet of a relay. When light shines on the cell, the current produced is strong enough to operate the electromagnet and close the switch of an electrical device; when light is removed from the cell, the current decreases to zero, and the switch on the electromagnet is pulled open, usually by a spring. This kind of arrangement makes it possible to operate automatically electric motors or other electric instruments that do such things as count moving objects, open doors, and operate burglar alarms merely when a beam of light is interrupted. In fact any operation that will interrupt a beam of light can be made to open or close an electric switch through the medium of the photoelectric cell.

For some types of work the photoelectric cell is much more satisfactory than human eyes, because it "sees" better than the eye and knows no fatigue. Since the current in a photoelectric cell is directly proportional to the amount of light falling on the cathode, it may be used for measuring the intensity of illumination. The foot-candle meters employed by lighting experts, as well as the exposure meters that are now a part of every photographer's equipment, are based on the photoelectric principle. It is possible, moreover, to arrange a photoelectric cell so that it will automatically control the amount of artificial light in a room by turning lights on and off. In this way the illumination is kept constant regardless of variations in daylight.

Special types of photoelectric equipment can be built for distinguishing different colors. One of the most elaborate is the Hardy photometer, designed by Prof. A. C. Hardy of Massachusetts Institute of Technology, which will analyze the color of any object when a sample of the colored material, such as a dyed fabric or a tinted paper, is placed in the machine. It records the wave lengths in angstrom units of every color reflected from the object and measures the relative strength

of each individual wave length. This feat would be totally impossible for the best trained human eyes to perform.

None of the uses of photoelectric cells just mentioned is so extensive as their application in the fields of taking motion pictures and television. Someone has said that "the heart of every sound motion-picture projector is a photoelectric cell." The same statement may be applied with even greater appropriateness to the television camera. Let us note briefly the part played by the photoelectric cell in these two important developments.

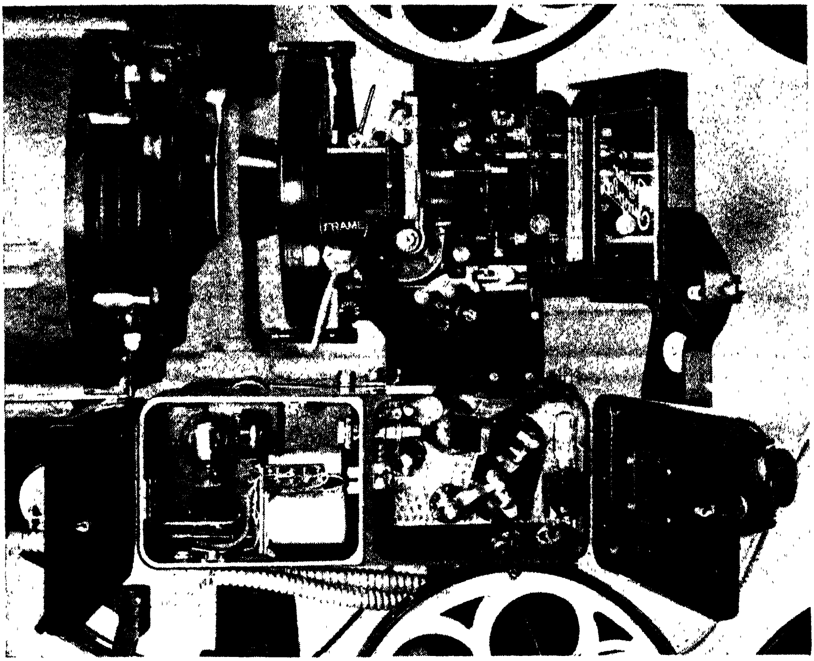
Making Movies Talk

The practical realization of talking motion pictures has come within the last ten or fifteen years, but the developments that made it possible go back much further. Shortly after the invention of the phonograph Edison developed a method of operating it in conjunction with a motion-picture machine so that the action on the screen could be accompanied by sound. This system was never used extensively, because it had too many limitations. The volume of the old-style phonograph was too low for large audiences, and the playing time of each record was too short. Moreover, it was almost impossible to keep the sound on the record synchronized with the action on the film.

Between 1920 and 1925 the electric phonograph was developed to the point where it would supply a volume of sound adequate for large theaters. Large records had also been perfected so that the sound accompaniment for a full reel of film could be recorded on a single disk. With these improvements began a large-scale development of talking-motion-picture equipment. The method was to run the motion-picture projector and the phonograph in exact synchronism so that each spoken word coming from the record would correspond to the lip motions of the speaker on the screen. This was an exceedingly difficult accomplishment, and everyone who remembers the early "talkies" will also remember that sometimes the speech got out of step with the actors, thereby creating some very amusing situations.

A practical solution to this problem of synchronization was provided by the photoelectric cell. The sound to be repro-

Motion-picture films with sound tracks. Variable-density track shown on film strip at (a) with magnified view at (b). Variable-width track shown on film strips at (c) with magnified section at (d).



Close-up showing interior of sound-motion-picture projector. Picture is flashed on screen by upper mechanism near control marked "Frame." Device for reproducing sound portion of film is contained in rectangular housing below. The incandescent lamp at extreme right projects image of sound track through lens system into photoelectric cell in circular glass bulb at extreme left. (Courtesy of Electrical Research Products, Inc.)

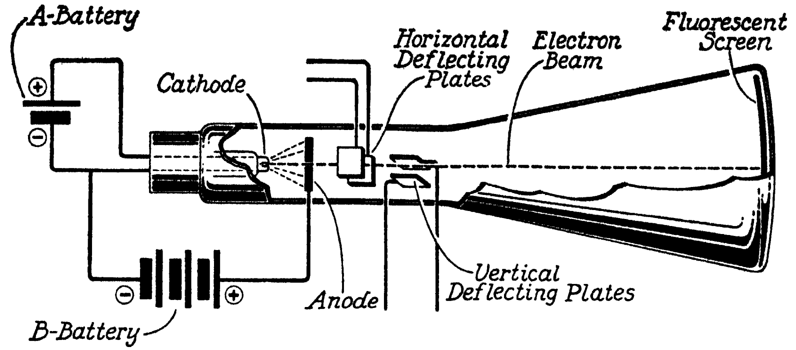
duced is recorded photographically along the edge of the film in a narrow strip called the sound track. This record consists of a series of light and dark lines. The spacing of the lines is proportional to the frequency of the sound, and their density varies according to its volume, or loudness. A beam of light from a lamp shines through the sound track upon a photoelectric cell. As the film moves downward, the light shining on the cathode of the photoelectric cell flickers in exact correspondence to the sound-track record. The flickering light produces a flickering electric current in the wires connected to the cell. This pulsating current is then amplified, and sound of corresponding frequencies is reproduced by loud-speakers placed behind the motion-picture screen, in a manner similar to that described earlier in the chapter for a public-address system.

The photoelectric cell makes it possible to use a record of the sound recorded on the film adjacent to the picture, and this is the basis of all modern talking-motion-picture systems. Using such a method for reproducing the sound makes it impossible, of course, for the sound to get out of synchronization with the pictures. In addition, many improvements have been made during the last few years in both the recording and the reproducing of sound on these sound-on-film records. The result is that the best talking pictures of today have a high degree of naturalness.

Making Electron Beams Visible

By this time you no doubt realize that the fundamental principle of all electron tubes is that the beam of electrons in the tube can be controlled by positively and negatively charged electrodes. Thus it is possible to make electron beams go through complicated motions at very high rates of speed. If we could see the beams in motion, we should soon be convinced of this fact. The feat has been partially accomplished by a special type of electronic device known as the cathode-ray tube. Electron beams from any electron tube are often called cathode rays because they come from the cathode of the tube. The special feature about the so-called "cathode-ray" tube is that the end of the tube contains a fluorescent screen which gives off a brilliant light when electrons strike against it, thereby making it possible to trace any motions of the beam across the end of the tube.

Perhaps an understanding of how the cathode-ray tube makes the movements of the electron beam visible may be gained from an analytical consideration of the accompanying drawing. The tube contains a heated cathode to produce the electrons. The anode immediately in front of it is designed with a small opening in its center. As the electrons are attracted toward the anode, some of them strike this hole and pass through the metal plate, and the rest are stopped by the anode. Those which do pass through the opening constitute a small beam of electrons. The electrons in this beam are moving at a high velocity, and they shoot on out to the end of the tube where they strike against it. The end of the tube is coated with a



In the cathode-ray tube, the electron beam may be deflected either horizontally or vertically so that it "draws" a picture or pattern on the fluorescent screen.

fluorescent mineral so that the electron impact produces a bright spot of visible light. Thus the exact position of the electron beam is shown by this light spot.

The other essential parts of the tube consist of two sets of deflecting plates placed at right angles to each other just beyond the anode, and mounted in such a manner that the electron beam passes between each set of plates. An alternating voltage applied between the vertical set of plates will move the stream of electrons back and forth across the end of the tube in a horizontal direction. This is accomplished by forces of attraction and repulsion; *i.e.*, at one instant one of the plates will have a positive charge, and the other will have a negative charge. At this instant the negative cathode-ray beam is deflected in the direction of the positive plate. An instant later the charges on the two plates have been reversed as the alternating current changes its direction of motion, and the cathode-ray beam is deflected in the opposite direction. The oscillating cathode ray produces on the screen a horizontal line of fluorescent light. Similarly, an alternating voltage applied across the horizontal set of plates will cause the beam to trace a vertical line of light on the end of the tube. By simultaneously combining the action of separate voltages on the two sets of plates all sorts of complex patterns may be produced.

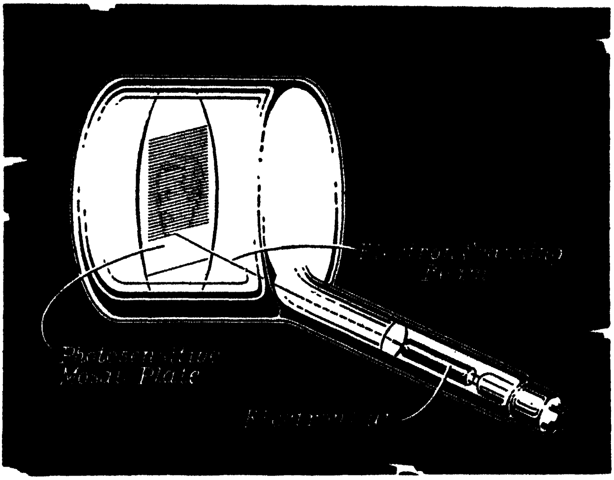
The cathode-ray tube is a very useful tool in studying the shapes of various forms of alternating-current waves. For such study, a slowly increasing positive voltage is placed across the

vertical plates so that the beam moves slowly across the screen in a horizontal plane. The alternating voltage to be studied is then placed on the horizontal plates. This moves the beam rapidly up and down in a vertical direction so that the final image on the screen is a stationary picture of the unknown wave. Using this method, it has been possible to make a study of any kind of wave motion that can be converted into an alternating current. When employed in this manner, the cathode-ray tube is called an oscillograph, because it "draws" a picture of an electric oscillation.

Broadcasting Moving Pictures

In 1884 the Japanese scientist Paul Nipkow invented a system for rapidly breaking up a picture into a number of units so that the units could be transmitted and again put together to form the complete picture. In 1932 the first television pictures were sent over radio waves from Alexandria Palace in London; and in 1939 regularly scheduled television programs of a few hours each week were broadcast from New York. These three dates are without doubt significant mileposts in the development of broadcasting moving pictures, which is what television really does. Television has reached a stage of technical development that gives excellent reproduction of pictures; also it is now a practical enterprise in this country, as it was in a number of European countries at the outbreak of hostilities in 1939. Reproduction of the pictures in a television receiver is accomplished by a cathode-ray tube which draws a moving picture of what another special kind of cathode-ray tube "sees" in the television camera at the transmitter.

From a technical standpoint television is a complicated and involved process; however, the basic steps are easy enough to understand. It consists of three essential processes: first, the breaking up of a picture into a large number of separate elements and converting them into electrical impulses; second, transmitting these impulses one at a time in sequence; and third, re-forming the elements in exactly the same sequence and converting them into a light image of the picture. The first step is accomplished by the television camera at the transmitter; the second is carried out over special telephone cables



The iconoscope of a television camera consists of a photosensitive plate on which is focused the image of the scene being televised by the camera lens at the right (not shown in the drawing) and an electron beam to scan the image.

or radio waves; the third is brought about by special receivers which assemble the electric impulses and convert them into light.

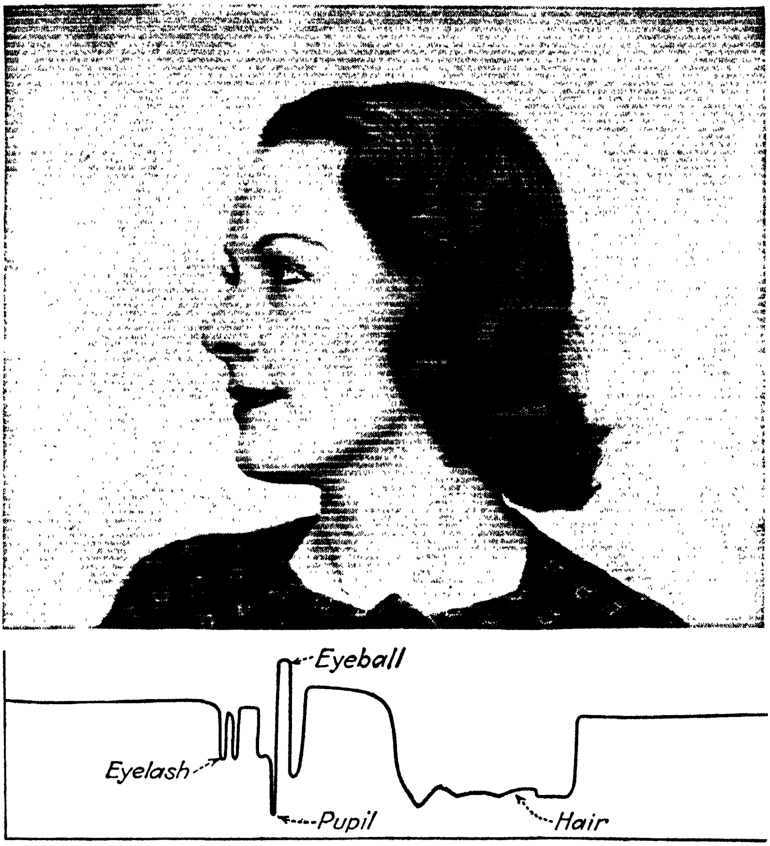
The television camera resembles an ordinary motion-picture camera in that it consists of a system of lenses for focusing on a "sensitive" plate an image of the scene being televised. But beyond this, any direct analogy with the camera is not found. The sensitive plate of the television camera is in reality a metal plate coated with an exceedingly large number of tiny photoelectric cells, or photoelectric particles. This photosensitive plate constitutes the image-plate end of a special type of cathode-ray tube, known as an iconoscope. When the light passing through the camera lens falls on the plate, it forms a visual image of the scene, just as it would be formed on any reflecting surface. In addition, the light also builds up an electric image of the scene. In the light areas of the image formed on the plate, negative electric charges are liberated by the photosensitive elements, and a positive charge is thereby formed on the image plate in exactly the same manner as a regular photoelectric cell operates. The strength of the positive charges will be in direct proportion to the strength of the light falling on the plate at the different points, and on areas of darkness in the image

no charges will be formed. In such manner, a positive electric image corresponding exactly to the light image of the scene is formed on the plate of the iconoscope.

The next thing that is done is in effect to break up this electric picture into a large number of separate elements, since there is no known way whereby the entire picture can be transmitted at one time. This is accomplished by a process known as "scanning" the picture. The meaning of this word in ordinary language is known to most people, and no doubt most of us have practiced scanning too much in the reading of textbooks.

Scanning a picture for television is essentially the same process as reading a printed page. It consists of breaking the picture up into a number of horizontal lines from top to bottom and then in effect separating each one of these lines into units of light and darkness, or, to be more accurate, into strong and weak electric charges in the electric image. These electric units correspond to the units of type in the line on the printed page. The process may be illustrated by drawing lines across a picture, as shown in the accompanying photograph. It will be observed that any particular line consists of different areas of light and darkness, and the electrical equivalent of one particular line is represented in the graph beneath the picture. This graph represents, therefore, the strength of the electric charges at different points in the particular horizontal line selected. Should these different patches of light and dark in each line be re-formed in their proper sequence in a receiver, and also all the other lines of the picture be arranged in the proper order from top to bottom, an image of the picture would be reconstructed.

In transmitting pictures by electrical means it is necessary to convert the separate areas of light and darkness into electric signals; hence the necessity of the plate covered with the photoelectric material to produce an electric image of the light picture. The transmission of the electric elements of the picture is a relatively simple process when a considerable length of time is available for rebuilding the picture. In transmitting a "still" picture the process can be slowed down as much as necessary, and the transmission and reception of the telephotos of still pictures has been a commercial process for many years. In actual television, on the other hand, the pictures to be



The scanning technique is represented by the parallel horizontal lines on the above image. The diagram below shows the values of brightness at points along the line that passes through the pupil of the left eye. (From Donald G. Fink, "Principles of Television Engineering.")

transmitted are likely to be of a moving scene in which the details are constantly shifting.

The scanning and transmission of these pictures must proceed at an extremely rapid rate so that no appreciable motion will occur in the scene during the transmission of any single view of it. Furthermore, in reconstructing any individual picture of the scene, the last element at the bottom of the picture must be reproduced before the first element at the top has faded from vision in order that the eye may see the whole picture at once. Fortunately for television, the eye does have a definite,

although very short, persistence of vision. The image formed in the eye of a fleeting scene does not fade the instant that the scene disappears but persists for a fraction of a second. There is no known way of measuring accurately the length of the persistence of vision, and it is known to change with a change in light intensity; it is, however, approximately one-thirtieth of a second. The complete television picture must be scanned, transmitted, and reproduced within this time therefore, for it is to be seen all at once. The standards established for modern television provide for reproduction at the rate of thirty pictures per second.

It probably can be understood without explanation that the fine detail of a picture and the fine degrees of shading can be reproduced only when a large number of separate elements of the picture are used in its reconstruction. Anyone who has ever observed carefully the number of dots used in printing photographs in newspapers, magazines, and books knows that the reproductions showing the finest detail are those having the largest number of dots, or separate units, other things being equal. The same thing holds true in reconstructing a television picture. The picture is scanned from top to bottom with 441 lines, this number being sufficient for them to blend together fairly well so as to be indistinguishable in the reproduced picture. If the picture is to have the same degree of resolution in the horizontal direction as it has in the vertical, each line must contain about 400 picture elements. This is the minimum standard set for modern television. It is achieved by having at least this many photosensitive elements in each horizontal line of the image plate in the iconoscope and a scanning beam small enough to cover each one of them separately.

A little calculation reveals the tremendous rate at which the picture is scanned and the enormous number of picture elements transmitted in one second. If the image scanned is 5 inches wide, the scanning beam must travel at the rate of $5 \times 441 \times 30 = 66{,}150$ inches, or about one mile per second. Actually the scanning beam moves at about twice this speed, since it must return to the opposite side of the image each time that a line is scanned. Assuming that each line consists of 400 picture elements, the system must be able to handle $400 \times 441 \times 30 = 5{,}292{,}000$ picture elements per second. This is the minimum

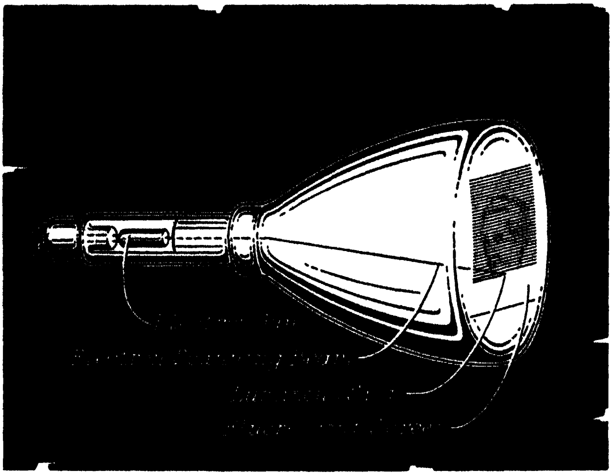
number for good definition, and rates as high as 8,000,000 picture elements per second are used in most modern television transmissions.

It is obvious that highly special equipment is necessary in order to get a scanning motion at a speed of one to two miles per second in both the camera and the receiver and that special transmission facilities are required to handle signals at the rate of several million per second. The only scanning mechanism that can be made to move with the desired speed is a beam of electrons, or the cathode ray. Cathode-ray tubes are used, then, in the camera to scan the image and produce a pulsating signal to be transmitted, and in the receiver to scan a fluorescent screen and convert the signals into a light picture.

The image plate in the iconoscope of the camera is scanned by a cathode ray from left to right and from top to bottom by directing, from a side arm in the tube, the cathode ray at the image plate, as shown in the drawing on page 471. As the electrons of the cathode ray fall on a point of high light intensity, its strong positive charge will be neutralized by the electrons, thus setting up a voltage change in the circuit connected with the image plate. As the cathode ray moves to a point of less light intensity, a weaker positive charge is encountered. This positive charge is neutralized, but a smaller voltage change in the plate circuit is produced. As the cathode ray moves across the image during the entire process of scanning, the voltage on the image plate continually changes in such a way that the change in voltage is proportional to the change in brightness of the picture elements.

In this manner the light and dark elements of the picture are translated into variations of the plate voltage and the plate current in the cathode-ray tube. The fluctuating plate current is then transmitted as the picture signal, or "video signal," either by wire or radio to a distant point. Radio waves are quite capable of handling a video signal that fluctuates with as many as 8,000,000 vibrations per second.

The television receiver is a device for converting the video signal into light and dark spots and assembling them in proper order for reproducing the picture. This is accomplished by a special type of cathode-ray tube, referred to as the "picture"



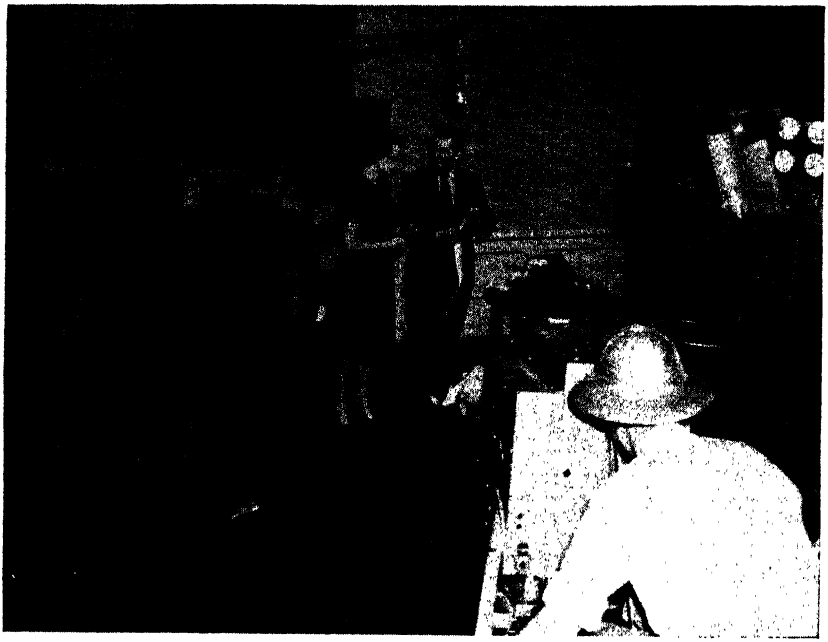
The kinescope of a television receiver consists of a cathode-ray tube in which the electron scanning beam varies in intensity so as to produce areas of light and darkness on the fluorescent screen in exact synchronism with the sending signal.

tube, or kinescope, a simplified drawing of which is shown. The motions of the cathode-ray beam in this tube are exactly synchronized with the motions of the beam in the transmitting tube. This means that the two beams are in precisely the same position at exactly the same instant. This accurate synchronization is accomplished by a separate signal which is sent from the transmitter to the receiver simultaneously with the picture signal.

The next step in reproducing the picture is to cause the video signal to vary the strength of the cathode ray in the receiving tube in exact proportion to the strength of the plate current in the sending tube. To accomplish this the video signal is fed into the receiving tube by means of a special grid. The effect of this grid is to cause the cathode-ray beam in the kinescope to vary in strength in exact proportions to the changing image-plate voltage in the iconoscope. As the cathode-ray beam gets stronger and weaker, it produces brighter and darker spots of light, respectively, on the fluorescent screen. While this occurs, the beam moves with lightning speed over every minute area of the screen. Each tiny section thus becomes either light or dark just as it was in the original picture, and the scene is reproduced on the fluorescent screen.

Actual televised reproduction of person shown in previous photograph. A 441-line image. (From Donald G. Fink, "Principles of Television Engineering.")

Television is an example of electron-tube application that has been developed to its present state as a result of painstaking research and experiments in a great number of laboratories. The work has been coordinated in such a way that television equipment is now fairly well standardized and receivers are available to the public. Although a wider use of television existed in Europe than in America prior to 1939, the technical development in this country was just as far advanced as in any foreign country. Television in foreign countries was a subsidized government monopoly, and its use was thereby accelerated for a time. In the United States the problem of full commercial television may be resolved into the following items: costs of broadcasting and research developments, nature and costs of programs, most suitable wave lengths for transmission, best scanning standards for favorable public reaction, and methods of transmitting programs over wide areas. Their solution involves the expenditure of large sums of money, and the dispatch with which they are solved will depend to some extent upon the public's reac-



New York University students (below) witnessing an educational television broadcast from the National Broadcasting Company television studio (above) in which science lectures illustrated by suitable working demonstrations were televised. (N. B. C. photographs.)

tion to television as a desirable form of entertainment and communication.

Progress in television broadcasting in the United States has been made since 1938, and many experimental types of programs have been initiated in the fields of entertainment, illustrated news, education, and national defense. These have included full-length shows from the theaters as well as special performances adapted to television, fashion parades, travelogues, baseball and football games, national conventions for the nomination of presidential candidates, and science lectures and demonstrations. The first use of this medium for educational purposes was made by Dr. C. C Clark in cooperation with the National Broadcasting Company in 1938 when a class of New York University students was given instruction in science subjects by means of television broadcasts from the studio to the classroom.

REFERENCES FOR MORE EXTENDED READING

Kaempffert, Waldemar: "Science Today and Tomorrow," Viking Press, Inc., New York.

The science editor of the *New York Times* has drawn upon his long experience as an interpreter of science to the public to write a book that contains a wealth of scientific information and his interpretation of what effects new discoveries are likely to have on individual and community life.

Mills, John: "Signals and Speech in Electrical Communication," Harcourt, Brace & Company, Inc., New York, 1934.

A well-known communications engineer has written an entertaining discussion of the basic principles in modern electrical communication, including television, that is highly recommended for the nontechnical reader.

Underhill, Charles R.: "Electrons at Work," McGraw-Hill Book Company, Inc., New York, 1933, Chaps. 9–26, inclusive.

The reader who wishes a simplified but accurate account of the fundamental principles that apply to modern vacuum-tube apparatus will find this a good reference book.

Hudson, Ralph Gorton: "Electronics," John Wiley & Sons, Inc., New York, 1932.

It is difficult to find a book that is a good connecting link between pure physics and applied electronics, but Mr. Hudson has written one. It is suitable for the reader who has a general knowledge of elementary physics but who has not heretofore given the subject of electronics much study.

SMITH, A. W.: "The Elements of Physics," McGraw-Hill Book Company, Inc., New York, 1938, Chaps. XLVI, XLVII.

Relatively few other textbooks of physics have such a good summary of the principles of modern vacuum tubes and photoelectric cells. A concise and accurate explanation for the student who understands the fundamentals of electricity.

NILSON, A. R., and J. L. HORNUNG: "Practical Radio Communication," 1st ed., McGraw-Hill Book Company, Inc., New York, 1935. Chaps. VII, VIII.

These chapters are devoted specifically to a description of the electroacoustic apparatus employed in modern radio-broadcasting studios and control rooms. Since radio and electronics are so closely related, the book is a good reference work on electronics.

DAVIS, A. H.: "Modern Acoustics," The Macmillan Company, New York, 1934, Chaps. V–VIII, inclusive, and XVII.

An advanced treatment of the electroacoustic apparatus used for the recording, reproduction, and measurement of sound.

HENNEY, KEITH: "Electron Tubes in Industry," McGraw-Hill Book Company, Inc., New York, 1936.

For anyone who wants the full constructional details on all types of vacuum-tube and photoelectric devices. The explanations do not involve any difficult mathematics, and the book is intended for the practical engineer and experimenter.

FINK, DONALD G.: "Engineering Electronics," McGraw-Hill Book Company, Inc., New York, 1938.

The student of general physics who wishes to learn something about the operation and design of electronic apparatus will find here a suitable treatment of theoretical principles and practical circuit operation.

FINK, DONALD G.: "Principles of Television Engineering," McGraw-Hill Book Company, Inc., New York, 1940.

A technical and very complete treatise on the subject of cathode-ray television. Recommended only for those who have a good understanding of radio and allied subjects.

Science Digest, published monthly by Science Digest, Inc., Chicago.

Articles include digests written for the layman of new science books, reports on scientific research, and science news.

Electronics, published by McGraw-Hill Publishing Company, Inc., New York.

A monthly magazine that contains both technical and semipopular articles on all phases of the subject. It is excellent for those desiring to keep in touch with the field through a minimum amount of reading.

Bell Telephone Laboratories.

15: MESSAGE'S WINGS

An Account of Electrical Methods of Communication

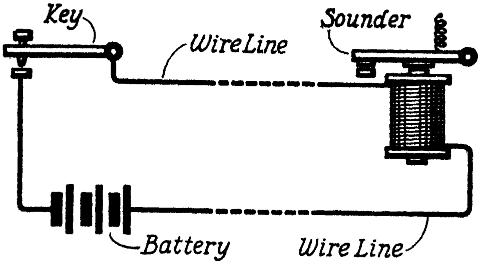
ON MAY 1, 1898, Andrew Rowan delivered his message to Garcia. The Spanish-American War had just been declared. President McKinley had ordered word sent to the Cuban General Juan Garcia, located somewhere in the mountain fastnesses of central Cuba, to ascertain what help he needed in his war against Spain. The mission was intrusted to Lieutenant Rowan, stationed then at Kingston, Jamaica. Within two days Rowan had crossed the Jamaican swamps, conscripted an open boat, and landed by night on the Cuban coast where he disappeared on foot into the jungles. Seven days later, after having narrowly escaped Spanish troops, he located Garcia at his mountain headquarters and delivered his message. His secret

journey did much to make the Spanish War a brief one, and for his heroism he received the Distinguished Service Cross. Since the time of this event, eighty million copies of an essay on his remarkable feat have been sold all over the world.

The communication system used by President McKinley to reach the rebel general employed the fundamental principle that had been used by man since before the time of recorded history; *i.e.*, verbal message was sent by a messenger. The ancient Sumerians and Egyptians as early as 5,000 years ago had extended such a communication system by installing relays of slave runners between cities. After the development of writing, regular caravans crossed the deserts carrying news and written messages; and later, ships on the sea and daring horsemen speeded communications over oceans and land.

From these early times to the middle of the nineteenth century, humanity had no other fundamental system of transmitting intelligence over long distances than to send a man with a message. Many methods for speeding his flight were developed, but man himself remained the carrier. This fundamental idea is still the basis of our postal system. Ships have been speeded until now they cross the Atlantic in less than four days; fast mail trains have replaced the Pony Express of earlier times; and now air-mail planes outdistance the trains. But man continues to be the physical medium for transmitting postal communications.

With the invention of the telegraph and a little later the telephone an entirely new system of communication was developed. Electric currents traveling over wires became the bearers of messages; no longer need man himself carry them. Electrical methods have reduced from days and weeks to minutes and seconds the time necessary to send transoceanic and transcontinental messages. At the turn of the present century a third fundamental communication system came into man's possession, that of sending messages by radio. Thus, messages are now transmitted at the speed of light, the highest attainable speed in the universe. Within recent years the beginnings of a fourth and fundamentally new system have been made, which is to send pictures by radio and to reproduce them with such speed that they may be viewed at the instant the event pictured is



Simplified telegraph circuit.

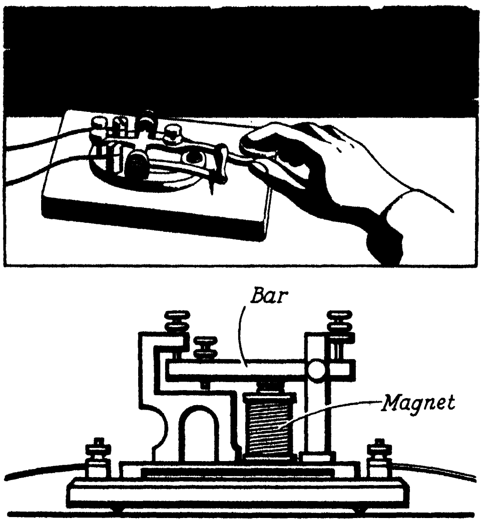
occurring. Television is now not only a technical accomplishment but also a commercial enterprise in some countries; and when vision is added universally to the telephone and radio, the last great frontier in communication will have been crossed.

Communication by Operating a Keying Switch

The practical application of electricity to communication began in 1844 with the invention by Samuel F. B. Morse of the telegraph, the essential principles of which are illustrated in the accompanying drawing. A simple switch, called a key, closes a battery circuit which contains a distant electromagnet. Above this magnet is mounted a small bar called a sounder which is held in the upper position by means of a coiled wire spring. As the key is moved up and down, it opens and closes the electric circuit through the electromagnet. When the circuit is closed, the sounder is suddenly pulled down against a metal stop so that it gives an audible click. The release of the key breaks the circuit and permits the spring to pull the sounder back up against another stop which makes a second click.

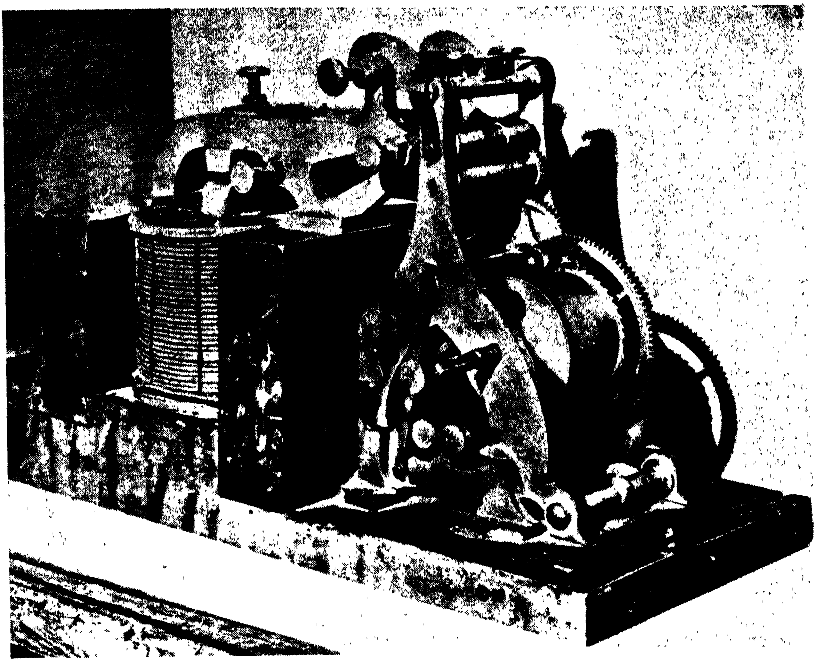
To send messages over such a system it is necessary, therefore, to have these clicks of the telegraph represent a form of alphabet so that words may be spelled out, transmitted, and received. The telegraphic alphabet consists of a code of dots and dashes. By depressing the key for only an instant, the operator transmits a dot to the sounder; a dash is transmitted by holding the key down for a longer time interval. In the regular sequence of dots and dashes, a short interval of no transmission corresponds to a space. Combinations of dots, dashes, and spaces represent the letters of the alphabet which afford a means of transmitting any kind of intelligence to a listener at the sounder.

The original telegraphic alphabet was invented by Morse and is known as the Morse code. As the years have passed, a modified alphabet known as the International Code has replaced to some extent the original Morse code. The original instrument that Morse used to pick up his message between cities differed from the modern sounder illustrated in the drawing in that the early sounder operated a pencil to record the dots and dashes on a revolving tape instead of making them audible, the tape being moved by a simple clockwork mechanism.



Depressing the key closes the telegraph circuit, and the telegraph sounder receives the dots and dashes as audible clicks.

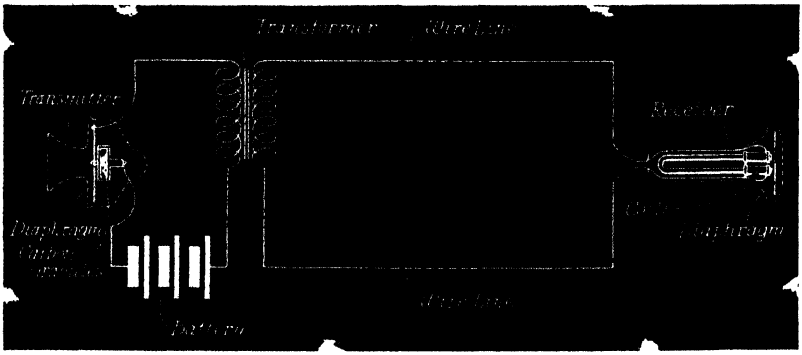
The chief limitation of the simple telegraph just discussed is that it cannot be used over very long distances. The resistance to the passage of electricity inherent in the wire line makes the signals of the sounder get weaker as the distance is increased, and beyond a certain definite limit it will not work at all. To correct this difficulty an electromagnet similar to the sounder is installed in the line at specified distances to operate a set of contacts instead of a sounding bar; this device is called a relay. The contacts close and open a second battery circuit so that the signal strength is renewed, and this process of rejuvenation of signal strength can be carried on indefinitely for transmitting signals over any desired distance. In practice, relays are usually used exclusively so that the key current never goes outside a local battery circuit in the telegraph station. Since sensitive relays can operate with less current than sounders can, the size of wire in the line may be greatly reduced, which produces considerable saving in the installation as well as better results in transmission.



The historic original Morse telegraph, sounder at the left and clock mechanism at right for operating a moving tape. (Science Service photograph.)

Modern telegraphy involves much more complex apparatus than the simple circuits described above. One important difference is that a number of messages are sent over one pair of wires simultaneously and in both directions. Such an arrangement is called the multiplex system of telegraphy. There are a number of methods for accomplishing this result, and in any case it is the usual practice to have the system operated by machine senders and receivers so that the operation is automatic. Both the messages being received and those being sent are recorded on paper strips called tapes. A series of alternating currents of different frequency usually carry the individual messages over the wires simultaneously. These different frequencies are separated at the receiving end by combinations of condensers and inductance coils, called electric filters, so that each message may be recorded correctly and separately from the others.

Although newer methods of communication have been invented and placed in use, the telegraph still holds its own as an



Simplified telephone circuit.

important system. Most of the business transactions taking place between the cities in a country are handled by telegraph when quick action and exact wording of a message are a necessity. Like the postal system, both the sender and the person to whom the message is addressed may have a copy of the exact message which can be filed and about which there can be no misunderstanding. The chief advantage of the telegraph is that it is much speedier than the postal system.

Communication by Voice Transmission

A typical instance that manifests the significance and distinction of the telephone as contrasted with the telegraph as a means of communication is shown by the messages of a salesman on a business trip. Referring to an important order or a cancellation, he would probably send to his business office a telegram, but he would use the telephone to talk to his loved ones at home. Herein is found one reason why the telephone has grown to such a gigantic system in modern life. It does more than the postal system or the telegraph; it allows instantaneous and personal two-way communication by voice. Even the tones of the voice and the way in which the words are spoken contain meanings that could not be transmitted by written words alone. This remarkable system, invented by Alexander Graham Bell in 1876, is an electrical means of transmitting speech that involves two essential elements. The first is the transmitter, or microphone; and the second is the receiver, or earphone. These two elements are, of course, connected by an electric circuit. A

simple telephone circuit is illustrated in the accompanying diagram. Upon examining this system, we see that it consists essentially of a transmitter connected in series with a battery and with the primary of a step-up transformer. The secondary of this transformer is connected to the receiver through an intervening line. The transformer is not altogether essential to the telephone operation, but the signals can be transmitted more efficiently when it is used.

Now let us investigate how the system works. Sound waves from the human voice, or other sources, moving through the air fall on the diaphragm of the transmitter and set it into mechanical vibration. The vibrating diaphragm will have the same frequencies, within certain limits, as the sound waves falling on it. As shown in the drawing, the diaphragm is arranged in such fashion that the vibrations alternately compress and extend one side of the box of carbon granules to which it is attached. This box of carbon is connected in series with the battery and the primary of the transformer so that the battery current must flow through the granules. When the granules are compressed, the current flows through with greater ease; but when they are extended, it meets greater resistance. The vibrating carbon granules, therefore, act like a variable resistance which causes the battery current to fluctuate up and down, and these fluctuations will be in exact synchronism with the variations in the sound waves striking the diaphragm. As a result, a fluctuating voice current with the same frequencies as the sound waves is set up in the microphone-transformer line and an alternating current with identical frequencies is transmitted from the transformer to the line.

These alternating currents pass over the line to the receiver where they are converted back into sound that reproduces the original spoken word. The telephone receiver consists of a U-shaped electromagnet with a steel diaphragm placed across the top of the U near the poles of the magnet. The alternating voice current flowing through the electromagnet produces around it a fluctuating magnetic field. Such a fluctuating magnetic field attracts and repels the steel diaphragm, thus causing it to vibrate with the same frequencies as the alternating current coming from the transmitter. This vibration of the diaphragm

Alexander Graham Bell at the opening of the New York-Chicago telephone line in 1892. (Courtesy of Bell Telephone Laboratories.)

in turn generates sound waves in the air around the receiver, the sound signals corresponding to those originally spoken into the transmitter. Such a circuit provides a means for the transmission of intelligible speech by electricity flowing along wires.

Modern telephony, like modern telegraphy, employs one circuit to transmit many messages simultaneously. In this case a number of alternating "carrier currents" of different high frequencies are sent over a telephone cable, one carrier current for each message to be transmitted. At the receiving end, each carrier current is filtered out by means of electric filters into separate channels where the original modulating signal is reproduced and transmitted to the local subscriber. Thus signals are not mixed, and a great saving in wire for long-distance lines is accomplished. In long-distance wire telephony, vacuum-tube

amplifiers are used to relay the signals that have been weakened by transmission over a length of line. These amplifiers, installed in central offices at definite points along the line, reamplify the signals so that they can be transmitted with full strength to almost unlimited distance on land or wherever wires may be strung and maintained.

At the New York World's Fair in 1940 visitors were given an opportunity to see these instruments functioning and to listen to a lecture explaining in detail how they worked. It is, indeed, a tribute to the invention of Bell that this exhibit was among the most popular of all demonstrations of a scientific nature. The attendance at the exhibit not only was a manifestation of the layman's interest in the details of operation of the telephone but also is an indication that this system of communication has become a necessity in modern life.

Radio Communication

Finally, we must turn our attention to that system of communication which flings the spoken word and music around the world with almost magical speed. During the last few years most people in the United States have sat listening to news reports from Europe's troubled scenes, broadcast direct to us by radio from the place where the events occur. The art of radio communication today shares, along with the newspaper, a greater ability to direct and control our changing civilization than does any other scientific development. This is true primarily because of radio's speed and directness. Its millions of code and telephone messages speed communications, news, and industrial operations and control armies and navies, shipping, and aircraft. Its direct appeal to the individual, through local broadcasting, and especially through controlled propaganda by various governments and other groups, is swaying and even changing the mode of thinking and living of great masses of the world's citizens.

Radio communication entails a much more complicated system than either the telegraph or the telephone. Let us consider first the general aspects of such a system and later discuss some of the finer points of its operation. In brief outline, the first step in the process is the generation of the radio waves.

The waves are then electrically modified or modulated in such manner as to conform to the signals to be transmitted. The modulated radio waves are broadcast into space, where they may be intercepted by a receiving set. Within the receiver the modifications, or signals, are removed from the waves, and they are eventually converted by the loud-speaker into the original sound signals. The methods whereby these steps are brought about are quite technical; however, the principles involved are not too difficult to be understood by the layman.

Modern radio was not developed by any single invention. Hundreds of improvements and developments have followed one another so rapidly that few people remember the many early experimental steps that were required to provide the radio of today. Much the same as was the case with wire communication, radio was first developed to send code telegraphic messages; then later it was discovered how to add voice signals to the radio waves.

As early as 1875 Heinrich Hertz had actually produced and experimented with very short radio waves. He produced them by electrical sparks in a spark gap connected to a miniature antenna and found that they could be sent out through space and received on a similar antenna. In these early days he had no vision of using them for communication. The work that he began was continued in 1894 by Sir Oliver Lodge, who contributed the idea of tuning, *i.e.*, developed a method of controlling the frequency, or wave length, of the waves. In 1896 Guglielmo Marconi, a former pupil of Hertz, continued the research and improved greatly upon the earlier techniques of generating, controlling, and receiving the radio waves. Marconi also used elevated antennas so that the waves might be transmitted over greater distances. But even with his own improvements, it was four years later before he was able to use radio for any sort of commercial communication.

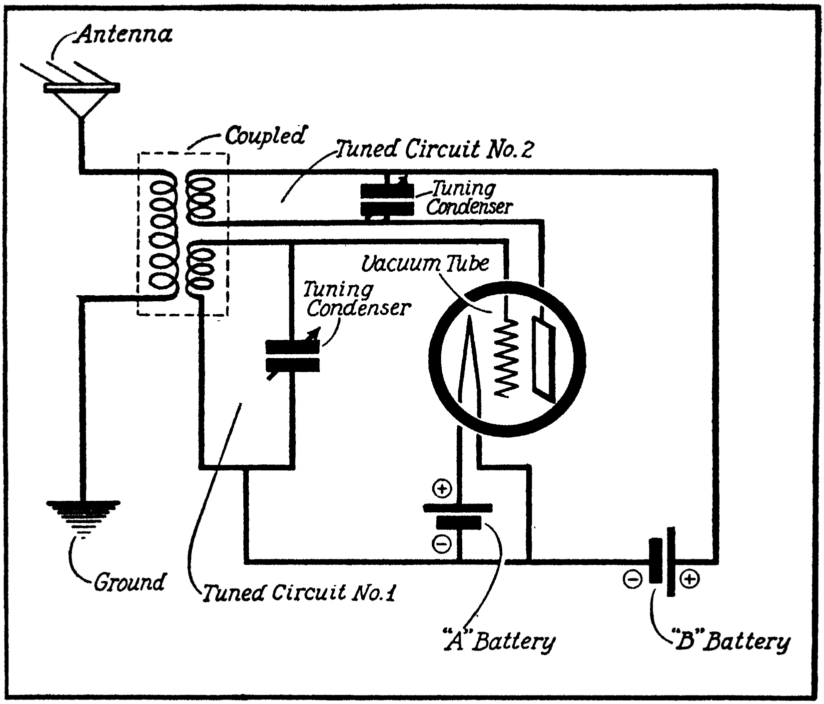
Up to the year 1914 radio communication was limited entirely to the transmission of code messages. Radio waves were generated by spark transmitters which produced broken or discontinuous waves, and the gaps in the waves were arranged so that they formed dots and dashes. These discontinuous oscillations were obtained by the use of electric sparks from high-

A picture of Marconi taken in the Cabot Memorial Tower, Bristol, England, in 1901. He is shown with the radio receiver with which he received the first radio message sent across the English Channel. (L. M. Cockaday photograph.)

voltage discharges and in a few cases by electric arcs not unlike the old-fashioned street arc lamp. During the first World War, however, a great change took place. It resulted from the invention of the vacuum tube and the momentous discovery that such tubes could be used to generate continuous electric oscillations for energizing radio transmitting antennas, which in turn produced continuous radio waves. This opened up the whole present field of communication by radio telegraphy, using a code of dots and dashes, and by radio telephony, using the spoken word in broadcasting. The vacuum-tube oscillator is, therefore, the essential feature of every radio broadcasting station desiring to transmit code, words, or music.

The Vacuum-tube Oscillator

It has been noted in a previous chapter that radio waves are produced by moving large masses of electrons back and forth in an antenna. These back-and-forth movements are the electric oscillations, or carrier currents, produced by vacuum-tube oscillators and referred to so freely in the technical language of radio. The invisible radio waves are formed in space around the antenna by these oscillations. Such waves are known as radio-frequency carrier waves, and they travel from an antenna in all directions. It is now possible to control the frequency of these carrier currents in the antenna and, thereby, the frequency and wave length of the radio carrier waves. The frequencies now



A vacuum-tube oscillator circuit for producing carrier currents in the antenna of a radio transmitting station. The two tuned circuits control the frequency of oscillation and, therefore, the wave length of the radio wave.

produced vary between 30,000 cycles, or vibrations, per second and 700 megacycles per second, a megacycle being 1,000,000 cycles.

The vacuum-tube oscillator consists essentially of two tuned electric circuits, one of which is used in the grid input (or "control") circuit of the vacuum tube, while the other is placed in the anode output (or "power") circuit of this same vacuum tube. These two tuned electric circuits are coupled to each other, either electrically or magnetically, so that some of the electric energy from the output power circuit (represented as Tuned Circuit No. 2 in the accompanying drawing) is fed back to the grid input control circuit (shown as Tuned Circuit No. 1). If the coupling adjustment is correct, just enough alternating current is fed back to the input circuit (from the power circuit) to keep the whole oscillator circuit functioning continuously so

as to generate in the anode circuit a strong alternating current, which can be fed to the antenna. Carrier currents of any desired frequency in the antenna can be produced by varying the capacity of the condenser and the inductance of the coils of each circuit in such manner as to match that frequency. The antenna will then radiate radio waves of that frequency. A simple vacuum oscillator is shown connected to a transmitting antenna in the diagram. It may be noticed that the two tuned resonant circuits shown are physically disconnected but magnetically coupled, as indicated by the dotted line.

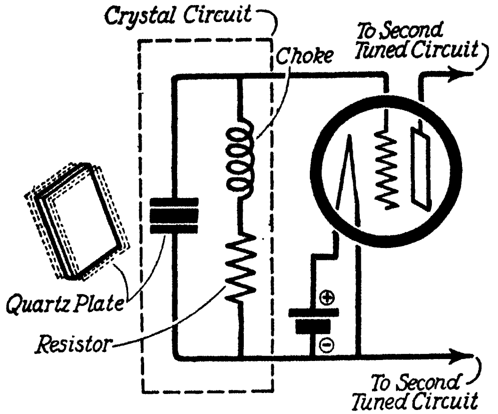
The Tuned Resonant Circuit

Tuned resonant electric circuits are used in radio whenever there is need for making a circuit respond to alternating electric currents of any particular frequency and not to any others. This includes nearly all radio receivers and transmitters. Any understanding of how a radio set is tuned to bring in one station and at the same time exclude others involves a general insight into the operation of a tuned resonant circuit. Such a circuit consists of an inductance coil with a tuning condenser connected across its terminals. Since a transmitting station is required by law to broadcast on one particular assigned frequency only, it must employ tuned resonant circuits that are carefully tuned to the station's assigned wave length. Likewise, receiving sets, in order to tune in different broadcast stations, must contain tuned resonant circuits capable of being adjusted so that any desired frequency may be received.

The relationship that exists between the frequency of an alternating current in a tuned circuit and the known value of the coil inductance and condenser capacity of that circuit is concisely and simply expressed in the formula

$$f = \frac{1}{2\pi\sqrt{LC}}.$$

In this formula, f is the frequency of the alternating current in cycles per second, L is the inductance of the coil in henrys, C is the capacity of the condenser in farads, and π is a constant the value of which should be known to every grade-school graduate.



A quartz plate "crystal" placed in the grid circuit of a vacuum-tube oscillator gives the oscillator greater frequency stability.

From the preceding relationship it will be seen that the frequency may be changed by varying either the inductance or the capacity or both. In most tuned resonant circuits, such as those used in radio transmitters and receivers today, the size of the inductance coil is of fixed value, whereas the capacity of the condenser is made variable. This arrangement reduces to one step the process of changing the frequency of the circuit, *viz.*, that of changing the capacity of the condenser. The frequency to which the set becomes resonant may be secured simply by rotating the control knob that turns the condenser plates, thereby varying their capacity. By such a process the set is "tuned" to a given station frequency.

When greater accuracy of tuning is required, especially in cases where a broadcasting transmitter must respond to one particular frequency allocated for its use and where another transmitting station has been assigned a frequency adjacent to it, a crystal-controlled oscillator is employed. If either one of these two transmitters should stray even slightly from its proper frequency, it would interfere with the other's transmissions. To prevent frequency drift, the crystal is inserted in the oscillator's grid circuit, replacing the first tuned circuit, as shown in the diagram.

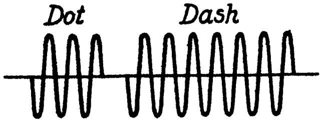
The crystal is a rectangle of quartz set between two conducting plates. It relies for its accuracy of tuning on the "piezoelectric" effect, which is, in brief, that the crystal changes its

shape when an electric potential is applied across its surface. If this potential is that of an alternating current, the shape of the crystal will change with each alternation, and the crystal will be set into mechanical vibration with a frequency that corresponds to the frequency of the alternating current. If the crystal is cut to a definite thickness, it will have a definite resonance, *i.e.*, it will respond to one definite frequency of mechanical vibration and to no other. When the crystal is ground thin enough, it will vibrate at a radio frequency, and its mechanical vibration period can be made to correspond to the electrical oscillation period occurring in the grid circuit of the vacuum-tube oscillator of a transmitter. In this way the frequency of both mechanical and electrical energies becomes locked into synchronism so that only the frequency equal to the natural one of the crystal can flow through the circuit, and the frequency of the transmitter is thus more finely controlled and stabilized.

Modern Radiotelegraphy

The output current of a crystal-controlled vacuum-tube oscillator will be, therefore, an alternating current of a constant and sustained frequency. Referred to in radio language as the carrier current, it is led into the antenna of the transmitting station, where the sustained oscillations in the antenna set up sustained and continuous radio waves of the same frequency which are transmitted into space much the same as light waves of given frequencies are transmitted from an electric-light bulb when the current is turned on. Such continuous radio waves are called carrier waves, and they are now used in modern radio telegraphy to transmit messages by the dots and dashes of the international code. This system of telegraphy is known as C.W., which means using continuous wave. The operation of C.W. telegraphy is rather simple in principle. When a carrier current flows in the antenna, a continuous radio wave of a corresponding frequency is transmitted; when it is turned off, no radio wave is sent out. By the manipulation of a telegraph key in the power circuit of the transmitter, it is possible to turn on and off the carrier current at intervals corresponding to the dots and dashes of the telegraph code. By pressing the key down for a short

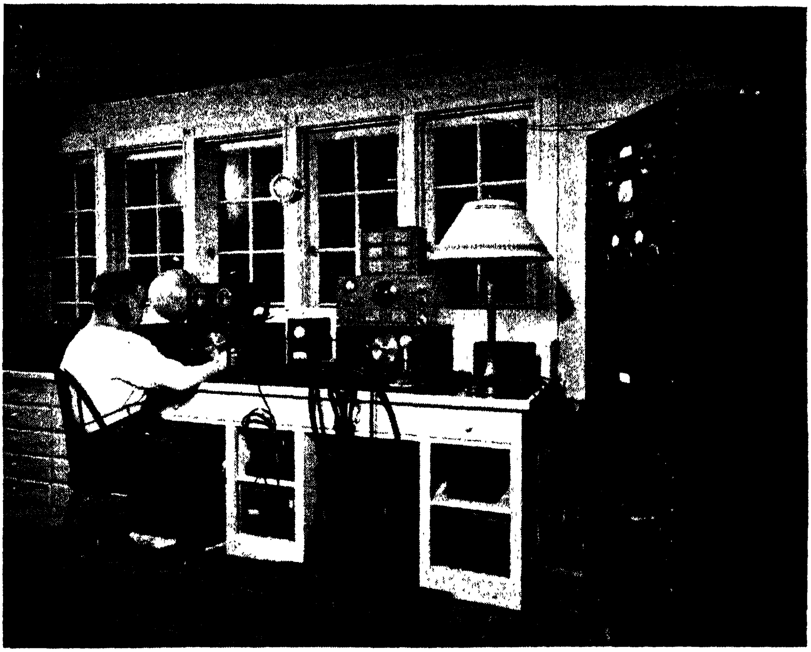
instant, the carrier is fed to the transmitting antenna, letting only a relatively small number of impulses through, and a "dot" is transmitted by radio waves. A longer depression of the key energizes the antenna for a longer period, letting more impulses through, and a "dash" is transmitted. Today many stations use automatic keying by tape-controlled senders, and hundreds of words a minute can be sent and received by commercial high-speed C.W. radiotelegraphy.



Telegraphy by means of radio waves.

Radiotelegraphy has one characteristic that makes its practical use more difficult than is obvious from such a simple explanation of the principles employed. The carrier currents induced in receiving sets tuned to C.W. telegraph transmitters are of a frequency that is too high pitched to be audible to the human ear. The radio receivers used for C.W. telegraph reception must have, therefore, additional apparatus that will make the dot-and-dash signals audible. This is what is known as a heterodyne. It consists of an oscillator that produces another carrier current of a somewhat different frequency from the carrier current received from the transmitter. The two carrier currents beat together, or heterodyne, to produce a signal of a lower frequency in accordance with the well-known principle of wave interference, discussed in Chap. 10. The secondary carrier current in the receiver is generated with such a frequency that it will heterodyne with the incoming carrier to produce a beat note in the range of audibility, and the dots and dashes may be heard. If, for example, the frequency difference between the two carrier currents is 500 cycles, a birdlike note is produced in the receiving head telephones or the loud-speaker whenever the key of the distant radiotelegraph transmitter is depressed to form dots and dashes.

Although C.W. radiotelegraphy uses the same general code method of sending messages as the line telegraph, it may be seen that it has no need of wires between the sending point and the receiving location. This makes it distinctly advantageous in instances where the transmission takes place over the ocean. There overhead wires are impossible, and underwater cables



Commander Cockaday tuning in a C.W. telegraph signal on a receiver containing a heterodyne oscillator to make the signal audible.

are costly to install and to maintain. Although the radio transmitting and receiving apparatus is much more expensive than the land-wire telegraph apparatus, the cost of year-in and year-out operation over long distances is greatly in favor of the radio system. The radio system may also be used between mobile units, such as ships, airplanes, trains, and automobiles, while in motion. This, of course, would not be possible with line telegraph.

How Antennas Send out Waves

As the antenna is the source of radio waves, it is important to consider just what an antenna is and a few of the principles upon which it operates. Most of us have come to think of an antenna as just a bit of wire strung around the room or out the window or on the roof. This is because we have used it only for receiving radio waves where its efficiency is of little importance, owing to the great amount of amplification in the modern receiving set. The transmitting antenna is quite a different story. For trans-

mitting purposes it must be as efficient as it can be made, since only a limited amount of energy can be derived from the transmitter for producing the radio waves.

The most efficient antennas are those which are properly tuned to the frequency being transmitted. The antenna is, of course, a part of the output circuit of the tuned resonant oscillator producing the carrier current, and its length is an important factor in the transmitter's efficiency in sending out radio waves at that frequency. In determining how long the wire of an antenna should be in order to be properly tuned, it is known that the wire has two electrical qualities, *viz.*, inductance and capacity, both of which affect the tuning and length of the wire. A short wire will have a small inductance and a small capacity; a longer one will have greater values of inductance and capacity.

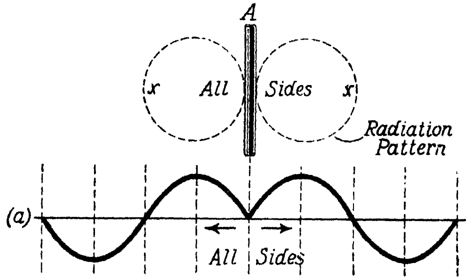
Let us examine the formula for calculating the wave length of a tuned circuit containing known values of inductance and capacity and see what this shows us about the length of the wire in an antenna. This formula is in reality the same one that was mentioned earlier in the chapter for determining the frequency of oscillation of an alternating-current circuit containing inductance and capacity. It is not too much to assume that the reader will remember from the discussion in Chap. 10 that there is a specific relationship between frequency, wave length, and velocity of the waves produced in any oscillating system. Making use of this relationship and substituting in the formula on page 493, we get the following statement:

$$\text{Wave length} = 300{,}000{,}000\ 2\pi\sqrt{LC}$$

where the 300,000,000 is the velocity of radio waves in meters per second. It will be noticed that L and C (inductance and capacity) are both multipliers; therefore long wires having large inductance and large capacity values will be resonant, or will tune, to a long wave length, and short wires having smaller values of inductance and capacity are resonant to shorter wave lengths.

From this formula a more specific relationship between length of the antenna and the wave length of the radio waves to be broadcast has been determined. Probably the most important point of this particular discussion is that it has been dis-

covered that the most efficient and the true-tuned antenna is one of such length as to be tuned to a half wave length of the radio waves being transmitted. An antenna of this type is called a half-wave antenna. The general reader should not interpret this to mean that it is one-half as long as the wave length of the transmitted wave, for that is not true; rather, it is of such a length as to be naturally tuned to one-half wave length of the radio wave.



A half-wave antenna A is shown at top with its radiation pattern, and the radio waves (a) leaving the antenna in all directions are represented at the bottom.

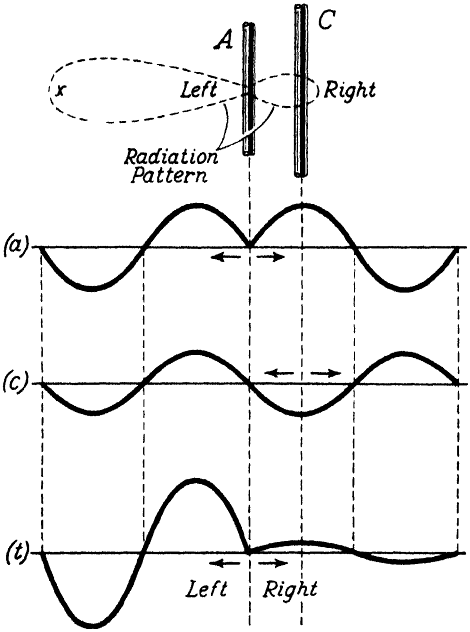
By making use of this discovery and by substituting the proper values in the preceding formula for wave length, a greatly simplified relationship has been worked out for determining just how long a half-wave antenna should be for any given frequency, and it is expressed as follows:

$$\text{Length of antenna in feet} = \frac{485}{f}$$

In this relationship f is the frequency in megacycles of the wave to be transmitted. As an example let us see how the relation works out for the proper length of an antenna for broadcasting on ten meters. A wave length of ten meters corresponds to a frequency of 30 megacycles, and 485 divided by 30 gives about 16.2 feet. Therefore, an antenna constructed with this length will be an efficient radiator of ten-meter radio waves.

A half-wave antenna mounted vertically will radiate waves in all horizontal directions, because the radiation pattern of an antenna is at right angles to the axis of the antenna. The accompanying diagram shows the circular lobes of energy on all sides of such an antenna, and at *a* are indicated the waves leaving the antenna in all directions. The radio-wave energy leaves the antenna in an expanding ring, filling space with radio waves in the form of a "doughnut" which is concentric with the axis of the antenna. The maximum power is radiated in the direc-

tion of points x, which are located along the extreme outer periphery of the doughnut-shaped radiation pattern. This type of antenna might be termed a general-coverage antenna. Broadcasting stations use this type, and often the antenna is a tall steel tower set up on an insulator at the base, in which case the whole tower acts as the radiator of radio waves.



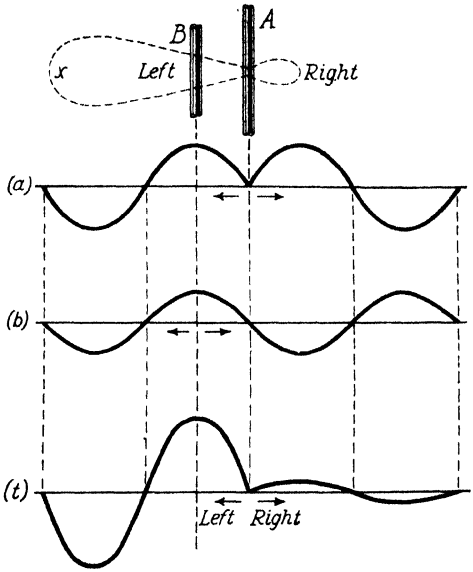
A two-element reflector-type beam antenna and its radiation pattern.

Beaming Radio Waves

It is often an advantage to have radio transmissions projected strongly in one direction with little or no energy going in others. Should there be a transmitting station in New York that maintained a regular schedule of broadcasting with a receiving station in Chicago, it would be a great advantage to send out most of the radiated energy in that direction only. This would be accomplished by transmitting in that direction what is called a "radio beam," produced by a special design and arrangement of the antenna.

Radio-beam antennas consist of a number of insulated rods, or wires, each of which radiates energy. The simplest beam antenna has two such rods, or elements as they are technically called. The principle of operation of the beam antenna may perhaps be made understandable by referring to the accompanying diagram. At the top is the main radiator A, which is connected directly to the transmitter. This radiator is a half-wave antenna and sends out radio waves in all directions, as just discussed. Its radiation wave is represented in the new diagram as a. As the radiation from element A passes to the right, it

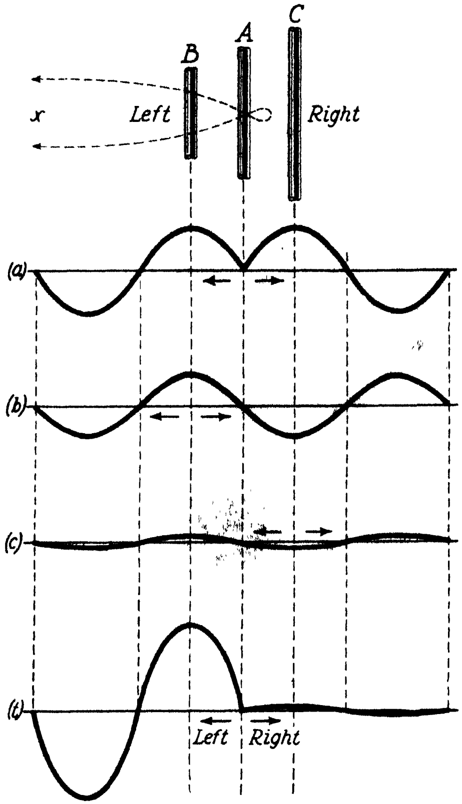
strikes the second element C, termed a "reflector," and it should be noticed that a reflector is greater in length than the main radiator. When the energy from element A strikes element C, a second carrier current is induced in C which causes C to begin to radiate wave energy as shown in the diagram as wave c. In this way two sets of radio waves are being sent out into space. Here is the important point: These two waves set up interference between one another. If the energy of a and c going toward the left is combined, shown as wave t, it can be seen that it increases the amplitude (and thus the total power) going in that direction. But if waves a and c are combined toward the right, it can be seen that wave c is opposed to wave a, and the result is only a tiny black wave. Therefore, the field pattern of this two-element beam has an extended loop to the left and a shortened loop to the right, shown in dotted lines; the front of the beam would be to the left, and it would send out almost three times as much energy in that direction as the ordinary half-wave antenna, whereas back radiation would be only a small fraction of that power. Thus a radio beam is produced.



A two-element director-type beam antenna and its radiation pattern.

Recently it has been found that by shortening the length of the second element of a two-element antenna, it could be changed into a "director"; *i.e.*, when placed near a radiator it would add to the strength of the combined waves in its own direction from the radiator rather than in the opposite direction, as is the case for the reflector type. In such an arrangement the second element is referred to as a director. Another diagram

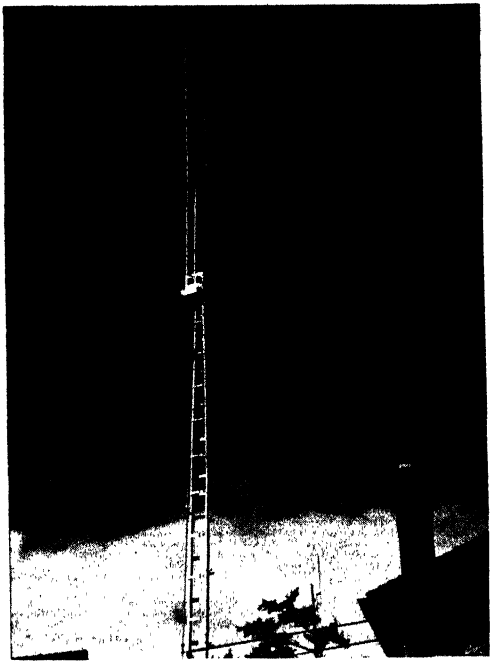
shows the position of the radiator A and the director B and the field pattern for this case. Also, the wave radiated by A and B and the total radiation in both directions are shown as a, b, and t, respectively. The director type of radio beam can be made to send out four to six times as much energy to the front as compared to the regular half-wave antenna's normal radiation.



A three-element reflector-director-type antenna and its radiation pattern.

By combining the principles of the director and the reflector into a single-beam antenna we have the three-element radio transmitter. This will increase the front radiation from ten to twenty times over the radiation from the half-wave antenna, according to the critical adjustment of the lengths of the three elements and the spacing between them. The accompanying diagram shows the method of combining the various wave energies a, b, and c from the three elements, the total being shown at t. Notice that the back radiation is almost eliminated but that the front radiation, to the left in the drawing, is enormously increased.

In summing up this study let us keep in mind that beam antennas consist of a radiator that is fed directly from the transmitter and also of one or more parasitically fed elements, such as a reflector or a director, used either separately or together to make up two-element or three-element antennas. Reflector elements are slightly longer than the radiator; and director



A two-element, vertical, director-type rotary-beam antenna. (L. M. Cockaday photograph.)

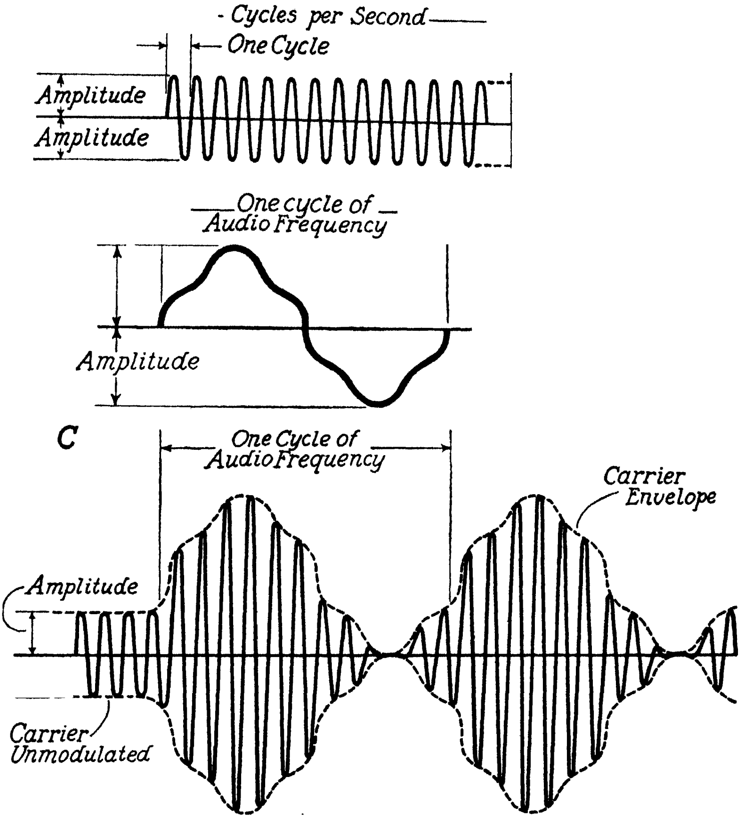
elements, slightly shorter. The reason for this difference in length, although too complicated for complete discussion here, is to change the phase of the currents generated in them so that they will either add their wave energy or subtract it in a given direction as the interference pattern is being formed. The final result of beam-antenna action may be likened to that of a searchlight where the rays of light are sent out in a narrow elliptical beam instead of radiating in all directions. Some radio-beam antennas are mounted on the top of a tall tower and are rotated by a remote-control motor so that they may be pointed in any direction, in this manner making it possible for a station to broadcast in any desired direction at will.

Modern Radiotelephony

As this picture of radio transmission begins to unfold, it should be perceived that transmission is accomplished by the radiation of a continuous carrier wave from the antenna. This

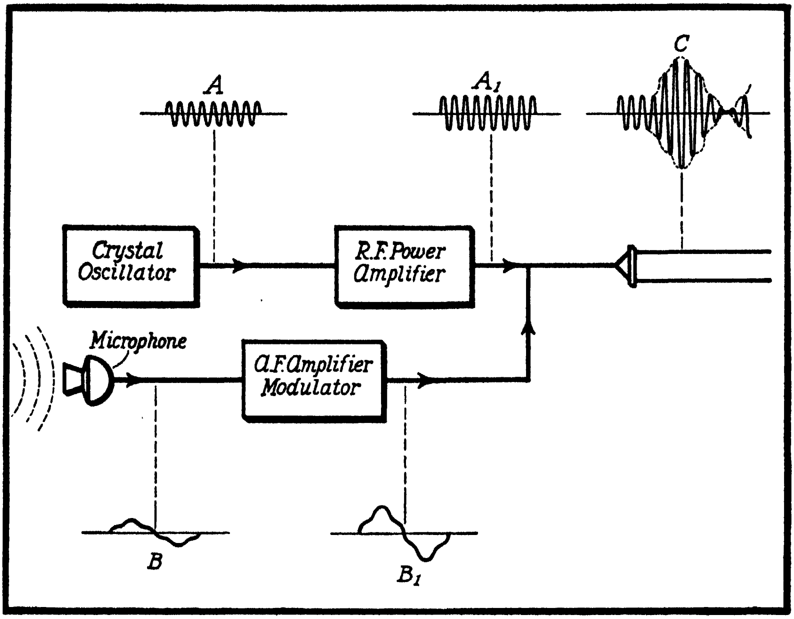
carrier wave is produced by the oscillating charges (or carrier current) in the antenna, and its frequency is determined by the crystal or tuned resonant circuits of the vacuum-tube oscillator in the transmitter. This is the fundamental way in which all radio transmission is accomplished. In radiotelegraphy the carrier wave is maintained at a constant amplitude, but it is stopped and started as desired to produce dots and dashes; in radio telephony, however, it is maintained all the time, and its amplitude is varied in strength with the frequencies of sound. When the carrier wave is maintained at a level carrier strength, no signal is heard in the radio receiver; but when a person speaks into the microphone, his voice controls the amplitude of the carrier wave, and a distant receiver, sensitive to changes in carrier-wave amplitude at an audio rate, reproduces the voice. This process of voice control of carrier-wave amplitude in a radiotelephone transmitter is known as amplitude modulation. Let us see how it works.

First, it is in order to study diagrammatically the three kinds of currents with which we are concerned in amplitude modulation. They are *A*, the unmodulated radio-frequency carrier current in an antenna; *B*, a voice current coming from a microphone; and *C*, an amplitude-modulated carrier current in the antenna. It is to be noted in the accompanying drawing that the unmodulated carrier wave is shown at *A*; and a voice current from a microphone, at *B*; particular note should be taken of the fact that the impressed wave, "or envelope," represented by the dotted line along the amplitude of *C*, has exactly the same wave form as the voice-current wave. It is further to be noted that the frequency of the modulated carrier *C* remains constant and the same as it is in *A*, as shown by the horizontal spacing of the solid carrier lines. Only the amplitude has been changed, so that at times its vertical excursions are increased or decreased in height in conformity with the voice-current wave form. It can be said that the wave form of the voice current has been impressed on the carrier current, so that the radio wave sent out from the antenna also carries with it the frequency of the human voice or of any other sounds produced before the microphone. Thus, the carrier wave has had its amplitude modulated by the voice current.



A radio-frequency carrier current flowing in a transmitter antenna has a constant frequency and amplitude, as shown in the top drawing. The center drawing is a diagram of a voice current. An amplitude-modulated carrier current in an antenna has a constant frequency but varies in amplitude according to the voice current shown in the lower diagram.

In order to examine the apparatus necessary to carry out this process of amplitude modulation in the radiotelephone transmitter, let us refer to the diagram on the following page. First there is the crystal-controlled vacuum-tube oscillator (represented at the upper left) which generates the carrier current A and supplies it to the radio-frequency power amplifier (shown at the center). The amplifier strengthens and supplies a greater carrier current $A1$ to the antenna (represented at the right). The circuit thus far considered has to do with the excitation of the antenna with the proper carrier current. The second circuit is concerned with the sound and modulation. It starts

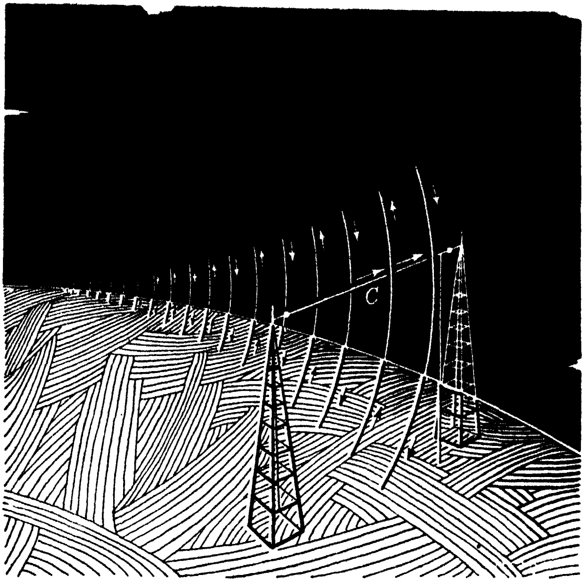


Block diagram showing the units of an amplitude-modulated radio transmitter and the currents in various parts of the circuit.

with the microphone (shown at the lower left), which converts the sound waves into an electrical voice current *B*. This current is supplied to the audio-frequency amplifier modulator where it is amplified to *B*1 and added to the anode energy of the radio-frequency power amplifier. The effect of adding this current is to increase and decrease the amplitude of the antenna carrier current *C* in conformity with the voice current *B*. Whenever the carrier-current level varies from the normal level by the action of the microphone and the modulator, the carrier radio waves sent out from the antenna carry the modulation as an envelope. Any radio receiver tuned to this frequency and picking up the modulated wave will reproduce the sounds falling on the microphone at the transmitter.

Practical Radio Reception of Amplitude-modulated Radiotelephony

It is to be hoped that a general understanding has been obtained from the foregoing discussion of just how radio carrier



Radio waves from a distant transmitter produce oscillating currents in a receiving antenna. A wave of one polarity produces a current C flowing in one direction in the antenna as shown by the two arrows, while the next wave of opposite polarity reverses the current in the antenna, and so on.

waves are generated, how they are modulated with waves corresponding to sound, and how they are transmitted into space. If it has, the next step is to consider how these waves are received in a radio receiver and how the original sound is reproduced. In taking up this discussion it should be kept in mind that radio reception is a step-by-step process which begins with the radio waves falling on the receiving antenna and ends with the sounds reproduced in the loud-speaker.

When radio waves strike an antenna, there is induced in that antenna a series of back-and-forth surges of electric charges, which flow through the lead-in wire to the input coil of the radio receiver. These oscillating charges produce a high-frequency alternating voltage in the coil, which corresponds exactly to the modulated carrier currents of the broadcasting station and has exactly the same frequency. This constitutes step one in radio reception.

The voltages induced in the receiving antenna differ from those producing the carrier current at the transmitter only in

that they are very much weaker, owing to the fact that the intensity of the radio waves diminishes in inverse proportion to the square of the distance from the transmitting antenna. For example, if the intensity of the radio wave is a certain value at one mile's distance, the intensity at two miles will be one-fourth; at five miles, one twenty-fifth; and at ten miles, only one-hundredth the strength. It can be seen, therefore, that at distances of thousands of miles the intensity of the waves may be so small that the voltage they produce in a receiving antenna may be of the order of only a few microvolts, a micro-volt being one-millionth of a volt. This represents small power, indeed, as it requires at least one volt at the detector tube to produce a satisfactory audible signal. It is exceedingly necessary for practical reception to have a means of strengthening, or amplifying, the received energy in order to use it to produce audible signals. This is especially true where loud-speakers are used. Amplifying the induced radio-frequency voltages is the second step in radio reception. There are two methods for doing this, and sometimes both are used in modern receivers.

One way to obtain radio-frequency amplification is by what is known as the "cascade" method. In this case, vacuum-tube amplifiers that operate at the radio-signal frequency are used. A tuning circuit is connected to the grid of a vacuum tube so that the tube amplifies the signal received from the antenna. Another tuned circuit of the same frequency is connected to the output of the first circuit in such a way that the amplified signal is impressed on to the grid of another vacuum tube which in turn amplifies it again. A number of these tuned-radio-frequency stages of amplification, one following another, may be used until the signal is brought up to sufficient intensity. All the amplifying is done at the same frequency; therefore, the tuning circuit in each stage has to be retuned when signals of another wave length are to be amplified for reception.

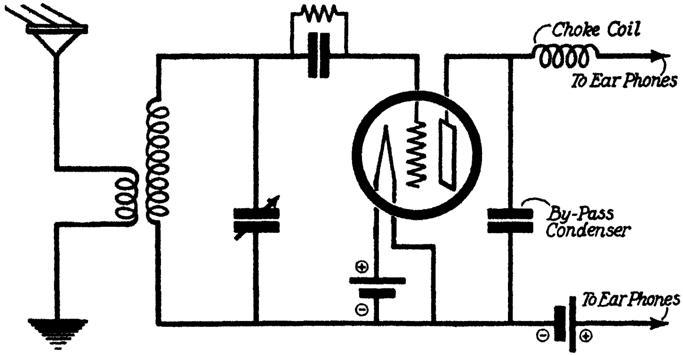
Another and far better method is the superheterodyne circuit which was developed several years ago by Maj. Edwin H. Armstrong when he was a signal officer of the United States Army during the first World War. He patented the invention and sold the exclusive rights to one manufacturer, so that for a considerable time it was impossible to buy a receiver designed

on this principle unless this particular manufacturer's set was purchased. This limited somewhat the wide use of the best general circuit in radio receivers ever invented. More recently all other manufacturers have been licensed to use this circuit, and today the superheterodyne method is universally used in the best radio receivers.

In the superheterodyne receiver a vacuum-tube oscillator is set so as to oscillate with a frequency that will produce beats with the incoming radio signal in such manner as to produce another but lower frequency signal. This signal is amplified at that chosen frequency by a number of vacuum-tube stages that are pretuned and fixed in frequency. In this method only the signal and the oscillator circuit have to be retuned when changing from one wave length to another, and usually both tuning condensers are connected to a single dial. This simplifies the design of the tuning circuits, while giving amplification factors of millions of times.

The received signals up to this point are radio-frequency impulses vibrating too fast to be heard by the human ear, even if led into earphones or a loud-speaker. In order to make them audible, the voice frequencies that form the modulated part of the radio signals must be separated from the radio-frequency carrier current. This process in the receiver of separating the audio frequencies from the radio frequencies is called demodulation, or "detection." It constitutes the next step in radio reception.

The modern method of detection used in radio reception employs a demodulating, or detector, tube, connected in one of a number of types of circuits, which separates the audio-frequency signals from the carrier current. One simple detector circuit is shown in the diagram herewith. The radio-frequency amplifier has been left out for simplicity, and the vacuum tube that serves as the detector is connected directly to the antenna and tuning circuit. A tuned circuit impresses the modulated carrier voltages on the grid of the tube, from which it passes into the anode circuit. This circuit immediately divides into two arms. One leads to the earphones, or loud-speaker amplifiers, and it contains a radio-frequency choke coil which stops the radio-frequency carrier current but allows the audio-frequency

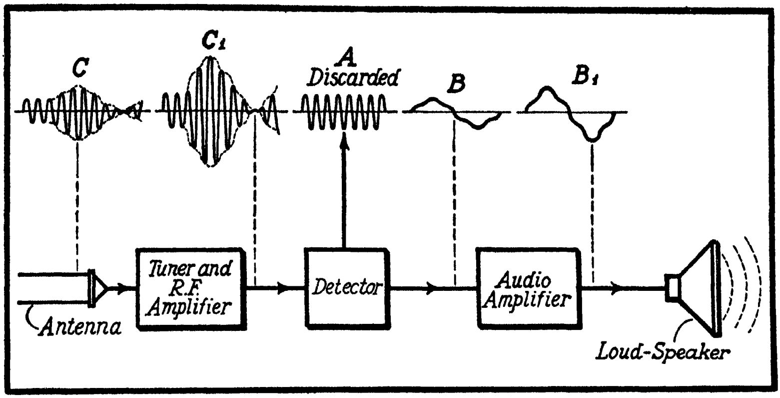


A detector circuit for rectifying or detecting the signal so that it can be made audible by headphones or loud-speaker.

current to pass. The other leads to the ground through the receiver chassis, and it contains a by-pass condenser which will prevent the audio-frequency current from passing but allows the radio-frequency carrier current to by-pass to the filament or cathode where it is neutralized. Thus, the carrier current and the modulated current come to the "parting of the ways," the first being by-passed, and the second continuing to be used in the loud-speaker after further amplification.

Amplifying the audio current and then converting it into sound in the loud-speaker constitute the last two steps in radio reception. They are the same as those described in the preceding chapter in connection with audio amplifiers and loud-speakers.

From this discussion it may be that the reader has arrived at the idea that radiotelephone reception is the reverse process of radiotelephone transmission, a conclusion that is in a general way correct. If we study the block diagram of the radio receiver, we may visualize the process already discussed. First, the receiving antenna has induced in it a weak modulated carrier current C as the result of excitation from the passing radio waves. This current is very low in amplitude, and the tuner and radio-frequency amplifier strengthen its amplitude as shown at $C1$. The current $C1$ is then fed into the detector where the radio-frequency carrier A is separated from the voice current B, which in turn is passed on to the audio-frequency amplifier. Here it is amplified and conducted into the loud-speaker, which converts the amplified voice current $B1$ into sound waves. Even



Block diagram of a radiotelephone receiver showing the various units and how the received currents progress and are changed until they energize the loud-speaker to reproduce the spoken word or music being transmitted.

a casual comparison of this block diagram with that for the transmitter will show that the two processes are similar but in the reverse order.

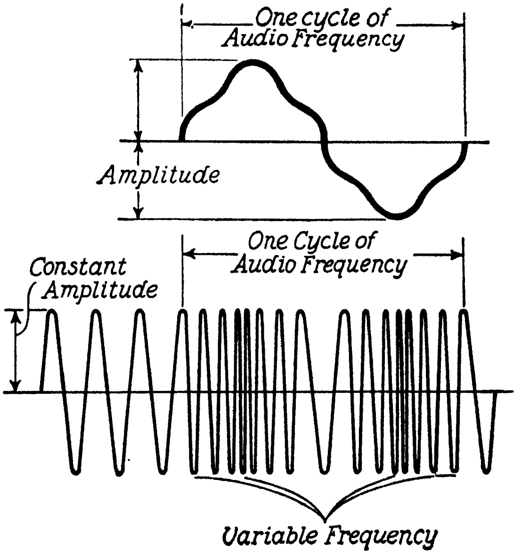
Frequency Modulation in Radiotelephony

A new system for transmitting audible programs that has recently had wide public acclaim and one that promises to become an exceedingly important factor in radio broadcasting utilizes another type of modulation. It is known as frequency modulation, and it is another contribution to the radio industry by Major Armstrong. This system is the realization of the dream of radio engineers for many years relative to two important aspects of radio, which are the elimination of static and the securing of high fidelity in the reproduction of music. Frequency-modulation radio produces noise-free reception under the most adverse conditions, and the lifelike qualities of sounds with a high range of frequencies are reproduced with such fidelity that listeners who are accustomed to ordinary radio are often startled by their naturalness. It was obvious that a system like this would not only create excitement in the radio industry but also would soon become of interest to the general public. Both these conditions have been fulfilled. The question most frequently asked by the layman is: What is frequency modulation, and how does it work?

It has previously been explained that commercial radio, as we have it today, is produced by amplitude modulation, whereby the carrier-current wave is varied in amplitude in accordance with the voice current's wave form. The frequency of the carrier current is kept extremely constant by specially designed crystal-control oscillators. The strength of the carrier current, represented by the amplitude of the wave, is increased and decreased at a rate that corresponds to the frequencies of the sound broadcast. In frequency modulation just the reverse of these processes is accomplished. The amplitude of the wave representing the carrier current is kept constant at all times, while the frequency of the carrier current is increased and decreased over a wide band of frequencies (approximately two hundred kilocycles) at a rate that corresponds to the frequencies of the sound being broadcast.

The two advantages of frequency modulation over amplitude modulation are inherent in the very principles operating in the two systems for the broadcasting of audio frequencies. In the first place, static disturbances are of such electrical nature as to produce amplitude variations in the radio waves. The amplitude-modulation receivers must of necessity respond to and amplify amplitude variations in order to reproduce the programs, and therefore they also respond to and amplify the static. There is theoretically no way to filter out the noise frequencies of static from the program frequencies except by eliminating most of the frequencies of the desired sound and thereby reducing the fidelity of its reproduction. Frequency modulation, on the other hand, is not troubled in this manner. Even though the amplitude of the frequency-modulated carrier wave is increased or decreased by the static, the receiver does not respond to the change of amplitude, since it is constructed to respond only to a change in frequency, and the noise is not reproduced in the loud-speaker.

In the second place, a band width of only twenty kilocycles at most is allowed in amplitude-modulation broadcasting, and this is not a sufficiently wide enough band of frequencies to permit (under practical conditions) all the frequencies of natural sound to be transmitted. Some of the high frequencies are eliminated, and in practice many of the higher frequencies as



A frequency-modulated carrier current varies in frequency in accordance with the voice current represented above, but the amplitude remains constant.

well as some of the low ones are cut off on the side bands. Frequency modulation is not troubled with this difficulty. Approximately 200 kilocycle bands is allowed for each station. This is a sufficiently wide range of frequencies to permit the transmission and reception of the very high-frequency harmonics of musical tones which give music its rich and brilliant qualities, as well as the low notes which give it smoothness and mellowness.

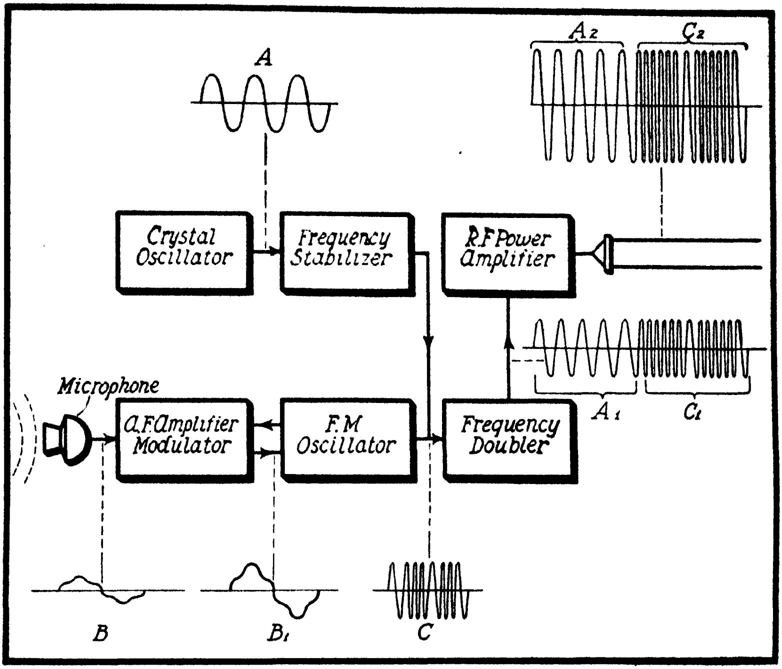
In order to obtain an insight into how frequency modulation works, let us examine the effect of a voice current on a carrier current produced in a frequency-modulation transmitter. In the accompanying drawing, the wave form of the voice current from the microphone is represented above, and the carrier current with its frequency modulated by this voice current is represented below. Notice that as the amplitude of the voice current increases, the frequency of the carrier current increases until the maximum amplitude of voice current is reached. Then as the amplitude of the voice current decreases, the frequency of the carrier current likewise decreases until the normal unmodulated frequency is reached. The same process is repeated

for the lower half of the voice-current wave. Particular attention is called to the fact that no change takes place in the amplitude of the carrier current.

Now look at the block diagram of the frequency-modulation transmitter shown in the next drawing. The circuit begins with a standard crystal oscillator which generates the carrier current *A* and furnishes it to a frequency doubler. This circuit, as the names suggest, doubles the frequency of the carrier current as shown at *A*1. The carrier current *A*1 excites the radio-frequency power amplifier, which supplies the amplified carrier current *A*2 to the antenna. From this it can be seen that the unmodulated carrier current is furnished in a manner similar to the amplitude-modulated transmitter, except that the frequency is doubled (for certain technical reasons which will not be mentioned in this discussion).

Next, let us consider the modulation system. It starts with the microphone which supplies the voice current *B* to the audio-frequency amplifier-modulator unit. This unit in turn furnishes a stronger voice current *B*1 to the frequency-modulation oscillator. This oscillator then takes control of the carrier-current frequency and changes it in accordance with the amplitude variations of the voice current, producing the frequency-modulated carrier current *C* which is supplied to the frequency doubler. The output of the doubler, shown at *C*1, is then supplied to the radio-frequency power amplifier, and its amplified output *C*2 in turn is applied to the antenna. It is to be seen, therefore, that when no modulation is present, *i.e.*, when no sound is entering the microphone, the crystal oscillator has full control over the final carrier-current frequency, as represented by *A*2; but when modulation starts, the frequency-modulation oscillator takes over control and varies the frequency in accordance with the frequencies of the sound falling on the microphone, as shown at *C*2.

The receiving apparatus for frequency modulation is quite similar to that for amplitude modulation, except that the detector circuit is one that responds to changes in frequency of the carrier rather than to changes in its amplitude. There is also a suppressor circuit which prevents any high-peaked amplitude pulses from being passed. The purpose and effect of the



Block diagram showing the units of a frequency-modulation radio transmitter and the currents in various parts of the circuit.

suppressor circuit are to eliminate the amplitude variations picked up by the radio waves from all sorts of static and electrical disturbances in the atmosphere and thus prevent them from reaching the loud-speaker and thereby producing the noise. The compressor has no effect, of course, on the audio frequencies from the studio, since they are represented by variations in the frequency of the carrier, which pass the suppressor and continue to the loud-speaker.

Uses of Radio in the World Today

Radio broadcasting has become one of the most common and at the same time one of the most spectacular uses of radio communication. Low-power stations of 25 to 1,000 watts serve small local communities in the United States with entertainment, news, and educational programs. Larger stations, a few of them with power up to 500,000 watts, serve larger areas with the best programs offered by radio, and they often are

interconnected in "chain" systems which cover the whole country. The local stations employ wave lengths ranging from 165 to 350 meters, and the wave lengths of the larger stations run from 300 to 550 meters. In the United States alone there are approximately six hundred radio stations of these types.

By far the most dominant note in commercial radio broadcasting in the United States at present is the operation of chain broadcasting by the large broadcasting companies. Chain broadcasting is accomplished by supplying simultaneously a group of broadcasting stations in different parts of the country with the same program. These programs may originate in one of a number of studios located in such cities as Chicago, New York, or San Francisco, and they are transmitted over special leased telephone lines so that each station may broadcast the program for its own community. The telephone-wire lines have been specially engineered to carry all the necessary frequencies of voice and music, so that the quality of the sound is not impaired by transmission of the signals over these lines. The chain system of broadcasting as used in America is exemplified by the Columbia, the Mutual, and the National Broadcasting systems.

The chain systems own, control permanently, or lease occasionally the broadcasting stations assigned to the systems and sell time on the air for advertising purposes. Fees collected for the advertising pay for the "sustaining," or "popular appeal," programs that the chains put on the air and for operation and maintenance of the stations, besides making a profit for the system. Each chain has a "key" station in important cities and other stations scattered throughout the network to pick up agricultural or other special programs. There are over three hundred stations in the United States that broadcast chain programs either full time or part time.

The American system of broadcasting is privately owned, but it is regulated by the Federal Communications Commission. This government commission also regulates public-service, telephone, telegraph, and cable systems. Its radio regulations and rulings are limited to allocations of licenses, time on the air, and frequencies to be used by the broadcasting stations. Regulations and rulings are authorized by Congress on the basis of

The first commercial radio station in the United States, Westinghouse KDKA in 1920. (L. M. Cockaday photograph.)

"public interest, convenience, and necessity." The commission may, by authority invested in it by Congress, set up and enforce regulations for program material to protect the public from "indecent, profane, or otherwise obnoxious content." However, no direct censorship of radio exists here as in broadcasting abroad.

Broadcasting has grown so tremendously during the last twenty years that it is now one of our largest industries and affects greatly the two other industries of radio, *viz.*, advertising and the manufacture and sale of radio-receiving sets for home use. The first actual broadcasting was done by amateurs who sent out music from phonograph records. A few of these amateur stations became the first licensed commercial stations. As an example we may mention the station of Frank Conrad which he built in his garage in Pittsburgh in 1920 and which he soon moved to the Westinghouse factory in Pittsburgh where he was employed. This station has now become the world-famous KDKA. Conrad's station when first assembled in the Westinghouse plant was located entirely in one room, with the trans-

mitter, the amplifiers, and microphone all grouped together. So great have been the popularity and growth of the station, along with the entire radio industry, that at present the large and powerful station requires a number of buildings to house the transmitters and power equipment, others for the studios and control rooms, and still others for editorial and administration offices. The history of KDKA has been duplicated in modified form by many other broadcasting stations. Most of the modern stations have extensive plants, a large personnel of technical experts for operation of the stations, and staffs of professional performers to provide some of the programs. It is not uncommon for a station in a large city to have its transmitter located miles away from the studios in a location that will provide the best transmission possible.

Broadcast programs in the United States can be divided into a number of classes. Entertainment leads all others in percentage of time on the air. Under the heading of entertainment are musical programs, both popular and classical; radio dramas; mixed programs of music and the sayings of comedians; amateur hours; and various types of quiz programs. A careful survey in 1940 shows that this type of broadcast material takes up about 60 per cent of the total time. The remainder is used about as follows: 6 per cent for agricultural, 1 per cent for fraternal, 8 per cent for civic, and 20 per cent for news programs.

A new broadcast band began to function in 1940 which may become one of the most popular and important yet in use. It is known as the Frequency Modulation Facsimile broadcast band. It occupies the radio spectrum between seven and one-half and six meters, or, to give its equivalent frequencies, from 40,000 to 50,000 kilocycles. It has been divided into a number of 200-kilocycle channels, one for each station. Full commercial rights have been granted the broadcasters who have received licenses for these stations, and they are entitled to broadcast regular commercial programs with the simultaneous broadcasting of visual programs of facsimile over the same wave lengths. A person owning a receiving set that will receive these programs may hear the program and at the same time see the printed material, such as pictures, maps, diagrams, and half-tone illustrations, sent to him as the program unfolds. This is the develop-

Sound and Facsimile programs being received. The picture is reproduced on a sheet of paper similar to a newspaper. (L. M. Cockaday photograph.)

ment which has been popularly referred to as the "talking radio newspaper"; technically it is termed "F.M. Sound and Facsimile Multiplexed."

Broadcasting in foreign countries is generally controlled and operated either by a monopoly set up by the government or directly by the government itself. There are a few exceptions, such as the broadcasting in South and Central America. Examples of government-sponsored monopolies are the British Broadcasting Company, the Canadian and Australian Broadcasting Companies, and the Japanese Broadcasting Corporation. Examples of government-operated systems are the German Rundfunk System and the Italian EIAR System.

Short-wave broadcasting stations now exist in most foreign countries as well as in the United States, and their programs can usually be heard around the world. They operate on groups of wave bands of approximately thirteen, sixteen, nineteen, twenty-five, thirty-one, and forty-nine meters. As a rule, the shorter wave lengths are better for long-distance transmission in the daytime, whereas the longer ones are better at night. The longer wave bands are also best suited for winter broadcasting,

and the shorter ones get better coverage during the summer months.

During the last five years the advent of the all-wave receiver has made almost all listeners familiar with short-wave broadcasts from overseas. Hearing London, Berlin, Madrid, Rome, and stations in South America, Africa, and even Asia is a daily occurrence for those who engage in long-distance listening as a hobby. News in English and other languages as well as large amounts of government propaganda, music, and talks by leaders in civic, professional, and business fields may be heard, day or night, from foreign countries when atmospheric conditions are favorable.

At the present time it is the custom to have these short-wave broadcasts of news from overseas arranged for in America by one of the broadcasting chains or picked up by the chain and rebroadcast over the regular chain stations. In carrying out such a procedure, a main receiving station, such as the one at Riverhead, Long Island, N. Y., receives the broadcast from overseas on one of the short-wave frequencies and retransmits it over telephone wires to the other broadcasting stations, where it is again transmitted locally on the regular commercial wave lengths. So rapid has been the development that we are likely to forget that rebroadcasts of foreign short-wave programs were a novelty only a few years ago. What a thrill many Americans felt to be able to hear the uncrowned King Edward VIII as he addressed the world at the time when he renounced the British throne in 1936 and publicly relinquished office in favor of his brother George! That was one of the first great rebroadcasts. Many, however, remember the earlier short-wave rebroadcasts from the Byrd expedition at the South Pole in 1934. In this case two-way messages were heard, sometimes rather garbled, it is true, as the short-wave stations at WGY in Schenectady, N. Y., and KDKA in Pittsburgh transmitted messages to the expedition. Today such a two-way transmission would be commonplace, all of which shows the progress made and the speed with which improvements are being made, applied, and accepted in public usage.

Another important use for radio telephony that has developed in recent years is two-way aviation communication. Such

Miniature personal radio transmitters and receivers, which can be carried in the pockets of clothing, have already been perfected. (Wide World photograph.)

communication is now considered an absolute necessity in connection with commercial, military, and naval aircraft navigation for receiving information as to weather conditions, dispatching, landing, emergency calls, and blind flying in foggy weather. Also, radio beams to guide planes along a given line of flight have been developed, and they now crisscross the country on the main airplane routes between cities and from coast to coast. A flier merely tunes his receiver to the radio-beam frequency and flies along it from one city to another. It is not necessary, perhaps, to do more than mention the necessity of radio in military aircraft maneuvering. Aviation and radio-beam transmissions may be picked up on wave-length bands near 1,000, 100, 60, 30, 15, and down as low as 2 meters.

As a result of amateur researches, portable radio transmitters and receivers are becoming increasingly popular in automobiles and small boats and for carrying around in the hand like a valise. It has been predicted that some time in the not very distant future every citizen may have reserved for him, somewhere in the little-investigated region below one

meter, a wave length on which he may use his own "personal" portable transmitter and receiver unit, a unit so small that he may carry it like a camera. Amateurs have built and operated many of these types of radio sets, so there is no doubt that they will work.

REFERENCES FOR MORE EXTENDED READING

BERNAL, J. D.: "The Social Function of Science," The Macmillan Company, New York, 1939.

A capable British author discusses thoroughly what science could do and what it does do in affecting social and economic conditions of mankind.

HARLOW, ALVIN F.: "Old Wires and New Waves," D. Appleton-Century Company, Inc., New York, 1936.

The early history of the telegraph, telephone, and radio is set forth in this attractively written and well-illustrated book. A minimum of scientific treatment is somewhat compensated for by the thoroughness of personal anecdotes.

DANIELIAN, N. R.: "A. T. and T.: The Story of Industrial Conquest," Vanguard Press, Inc., New York, 1939.

The author treats the great Bell Telephone System as a historical laboratory and uses it to make a study of the scientific, financial, and economic forces that have influenced American industrial development.

GLASGOW, R. S.: "Principles of Radio Engineering," McGraw-Hill Book Company, Inc., 1939.

Although the book is designed for the engineering student, the layman will be interested in a considerable amount of material on the fundamentals of radio communication and the application of these principles to radio broadcasting and reception.

RIDER, JOHN F., "F. M.," John F. Rider, Publisher, Inc., New York, 1940.

"F. M." is written for and dedicated to the radio service man; however, the first two chapters contain a well-presented and nontechnical explanation of the principles involved in this newest type of radio.

Radio and Television, published by Popular Book Corporation, Springfield, Mass.

A trade journal that contains occasionally an article on some phase of the radio industry that is of interest to the layman. The magazine consists essentially of short semitechnical articles, news items for amateurs, and advertisements of radio products.

The Bell Laboratories Record, published by Bell Telephone Laboratories, Inc., New York.

Published by the research organization of the telephone industry, this magazine contains professional, technical, and semipopular papers on a wide variety of topics relating to the characteristics and handling of sound.

INDEX

W

X

Y